

168
13.8.70

THE
QUARTERLY JOURNAL
OF
EXPERIMENTAL
PSYCHOLOGY

478J
Editorial Board

D. RUSSELL DAVIS

R. C. OLDFIELD

J. W. WHITFIELD

G. C. DREW


M. D. VERNON

G. C. GRINDLEY

O. L. ZANGWILL

CONTENTS

A. CARPENTER. A Comparison of the Influence of Handle Load and of Unfavourable Atmospheric Conditions on a Tracking Task - - - -	1
O. L. ZANGWILL. Amnesia and the Generic Image	7
J. C. TSAO. Studies in Spaced and Massed Learning -III. Paired-Associate and Serial Learning - -	13
A. W. HEIM and J. G. WALLACE. The Effects of Repeatedly Retesting the Same Group on the Same Intelligence Test: II. High Grade Mental Defectives	19
W. E. HICK. The Threshold for Sudden Changes in the Velocity of a Seen Object - - - -	33
Proceedings of the Experimental Psychology Group 1948-9 - - - - -	42

 PUBLISHED FOR
THE EXPERIMENTAL PSYCHOLOGY GROUP

BY

W. HEFFER & SONS LTD. CAMBRIDGE

Ednl. for Research
13. 8. 70
9. 166

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Editorial Board

D. RUSSELL DAVIS	
R. C. OLDFIELD	J. W. WHITFIELD
G. C. DREW	M. D. VERNON
G. C. GRINDLEY	O. L. ZANGWILL

CONTENTS

D. RUSSELL DAVIS and DURGANAND SINHA. The Effect of One Experience upon the Recall of Another - - - - -	43
E. N. WILLMER. The Relationship between Area and Threshold in the Central Fovea for Lights of Different Colours - - - - -	53
K. R. L. HALL and R. C. OLDFIELD. An Experi- mental Study on the Fitness of Signs to Words - -	60
D. E. BERLYNE. Stimulus Intensity and Attention in relation to Learning Theory - - - - -	71
J. W. WHITFIELD. The Imaginary Questionnaire -	76
J. SZAFRAN and A. T. WELFORD. On the Relation between Transfer and Difficulty of Initial Task - -	88
Book Review - - - - -	95
Experimental Psychology Group - - - - -	97

PUBLISHED FOR
THE EXPERIMENTAL PSYCHOLOGY GROUP
BY
W. HEFFER & SONS LTD. CAMBRIDGE

THE
QUARTERLY JOURNAL
OF
EXPERIMENTAL
PSYCHOLOGY

Editorial Board

D. RUSSELL DAVIS
R. C. OLDFIELD J. W. WHITFIELD
G. C. DREW M. D. VERNON
G. C. GRINDLEY O. L. ZANGWILL

CONTENTS

E. C. POULTON. Perceptual Anticipation and Reaction Time	99
L. WEISKRANTZ. Figural After-effects in Stroboscopic Motion	113
J. A. DEUTSCH. A New Phenomenon of Apparent Movement and its After-effect	119
R. G. HOPKINSON. The Multiple Criterion Technique of Subjective Appraisal	124
D. RUSSELL DAVIS and DURGANAND SINHA. The Influence of an Interpolated Experience upon Recognition	132
MISCELLANEA	
J. W. WHITFIELD. Note on the Correlation of Errors and Performance in a Test	138
A. CARPENTER. A Simple Class-room Chronoscope	140
Book Review	142

PUBLISHED FOR
THE EXPERIMENTAL PSYCHOLOGY GROUP
BY
W. HEFFER & SONS LTD. CAMBRIDGE

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Vol. II

1950

Part I

A COMPARISON OF THE INFLUENCE OF HANDLE LOAD AND OF UNFAVOURABLE ATMOSPHERIC CONDITIONS ON A TRACKING TASK

BY

A. CARPENTER

478 J

(From the Medical Research Council Applied Psychology Unit, Cambridge)

An experiment was made to determine the extent to which the performance of a tracking task was adversely influenced by simulated tropical environmental conditions, by comparing this effect with that of loading the control handle with weights. The pursuitmeter used was similar to that used by Mackworth (1948) when he showed that the accuracy of performance was reduced under tropical conditions. The score of error was the total angular separation between the target and follower pointers expressed in arbitrary units. It was found that as the room temperature was raised from approximately 80° to 90° on the Effective Temperature scale, this score increased in the ratio 1.13, an increase which was also obtained by increasing the load on the control handle from 8 to 30 lb. There were also indications that the effect of room temperature was greater with higher handle loads. The effect on performance was not correlated with rectal temperature or with the amount of weight lost.

I

INTRODUCTION

PREVIOUS experiments in the Cambridge hot-room, and elsewhere, have shown that in a variety of different tasks, performance becomes less good as the temperature of the room is raised above about 85°. The tasks in which this deterioration has been found have included visual and auditory vigilance (Mackworth, 1950), rapid repetitive highly-learned tasks such as the reception of telegraphy (Mackworth, 1946), and the setting and decoding of messages on a mock-up coding board, purely physical work, using an ergograph, mental work, involved in the tracing of a hidden electrical circuit (unpublished), and a tracking task such as was used in the following experiment.

In all these experiments deterioration in performance at higher temperatures has been found, to an extent which was statistically reliable, and in some cases an approximate assessment of the degree of impairment was possible. For example, in the telegraphy experiment, the errors increased from an average of 3.2 per 1,000 characters at 79° on the Effective Temperature scale, to 25.3 at 97°. The difficulty in attempting to measure the performance on synthetic tasks such as those provided by an ergograph, or the tracking task provided by the pursuitmeter in the following experiment lies mainly in understanding the measuring scale. The proportion of errors made in telegraphy reception is a meaningful measure of performance, but in such a highly artificial task as tracking on a pursuitmeter, the score obtainable on our machines—merely the total angular displacement between the target pointer

and the subject's pointer—is really only of value as a means of comparing one condition with another.

This present experiment is an attempt to give a little more meaning to the extent of deterioration of this tracking task, by comparing the effect of room temperature with that of loading the handle with known weights.

A second purpose in the experiment was to find out whether there was an appreciable difference in the effect of temperature when the load on the handle was varied, that is to say, to discover whether the effect of high temperatures was mainly on physical work, or on the more skilled side of tracking.

II

APPARATUS

The pursuitmeter used in this experiment consisted of a display of two pointers, of white ivory, about $\frac{3}{16}$ in. \times $\frac{1}{4}$ in., with a bold central indicator line marked in black ink. These moved along an arc of radius about 8 in., and up to 3 in. in length, the upper being moved irregularly to and fro by an irregular cam operated by an electric gramophone motor, and the lower being operated by the subject by a horizontal handle 2 ft. in length, which was moved up and down through a maximum distance of 18 in. On this handle lead weights could be hung, giving loads, as measured at the end held by the subjects, of 2, 8, 16, 24, 32 and 40 lb. The task was to follow the irregularly moving upper pointer with the lower, keeping the two in alignment.

III

PROCEDURE

Three such machines were used, and three subjects tested at a time. There were four groups of subjects, and each performed the test four times, once at each of four room temperatures, which were 85°/75°, 90°/80°, 95°/85°, 100°/90° (° F. dry bulb/° F. wet bulb), the air velocity being approximately 100 ft. per minute. These four conditions were arranged with respect to subject group and to serial order (called "days") in a Latin Square arrangement.

In any one day each subject did eighteen test spells, each of 3 minutes, which consisted of one spell on each machine with each of the six weights. Within this framework, conditions were randomized.

IV

SUBJECTS

The twelve subjects were physically fit young Naval ratings, aged between 20 and 23 years. They were "acclimatized" to hot-room conditions by spending 3 hours per day for the fortnight before the experiment in the hot-room, at a temperature of 100° F. dry bulb/90° F. wet bulb. They wore shorts only, while in the hot-room.

Rectal temperatures were taken (a) on entering the hot-room, (b) after 1 hour of acclimatization and (c) on completion of the experiment. Fluid lost as sweat during each experiment was assessed by keeping a balance sheet of body weight to the nearest 5 gm.

V

SCORING

The total angular separation, irrespective of sign, between the two pointers of each machine, which occurred during each 3-minute experimental spell, was recorded on a ratchet disc and taken as the error score for that spell.

VI

RESULTS

The experimental design enabled the five major sources of variation in the results obtained, to be separately evaluated; namely, that due to the use of different machines, that between subjects, the effect of practice, of room temperature conditions, and of handle load. Of these, only the last two are experimental variables of interest, and the purpose of separately evaluating the effects of the first three is to exclude the variance of these effects, and to prevent it from reducing the accuracy of the estimation of the effects of room temperature and of handle load. They will not, therefore, be given here.

Table I gives the total error scores obtained at each of the six degrees of handle load and of the four room temperature conditions.

TABLE I
ERROR SCORES

Room Temperature	Handle Loading in lb.						Totals
	2	8	16	24	32	40	
85/75	3,363	3,259	3,278	3,403	3,614	3,763	20,680
90/80	3,388	3,314	3,362	3,492	3,734	3,862	21,152
95/85	3,762	3,565	3,703	3,989	4,329	4,625	23,973
100/90	3,644	3,588	3,680	3,938	4,030	4,436	23,316
Totals	14,157	13,726	14,023	14,822	15,707	16,686	

(a) *Effect of Room Temperature.*

The temperature totals in Table I show a marked increase of error score between room temperatures of 90°/80° and 95°/85°. The standard error of these total figures, taken from an analysis of variance, is 350, and thus this increase has a critical ratio of 8.06 and is highly significant, although the differences between the two lower temperatures and that between the two higher temperatures is not significant.

This agrees very well with Mackworth's figure of 94/84 (° F. dry bulb/wet bulb) as the point at which deterioration in performance with temperature becomes statistically reliable.

An approximate quantitative estimate of the degree of this deterioration can be made by taking the difference between the mean scores at the two lower and the two higher temperature conditions. This gives an increase of error score, for an increase of 10° F. in temperature, in the ratio 1.13.

(b) *Effect of Handle Load.*

The variance ratio of the error totals due to the several handle loads is 23.05, and the standard error of any one total is 238. Thus the general trend of increase in error with load is significant, although the drop in error between loads of 2 lb. and of 8 lb. is not so. This drop, however, is interesting in view of Hick's finding (Hick, 1946) that moderate frictional load, in a tracking experiment with crank handles, improved the accuracy of performance, presumably by providing an influence stabilising speed. In the present case a medium weight of 8 lb. might improve performance by reducing rapid small correcting movements, which while not con-

tributing much to a truly integrated score, would be counted fully on a ratchet device which takes no account of duration of error.

If the ratio 1.13 found between the scores obtained at the higher and lower temperatures is compared with the effects of handle load, it will be seen that this same ratio is obtained between the scores at loads of 8 and 30 lb., or between 16 and 32 lb. Thus the effect on performance of increasing the room temperature from 87/77 to 97/87, i.e. by ten degrees, is approximately the same as that of adding a further 20 lb. weight to the load on the pursuitmeter handle.

(c) *Interaction of Load and Room Temperature.*

This experiment was designed to provide some evidence as to whether the effect of room temperature was greater with the heavier handles, or, to put it another way, whether the effect of handle load was greater at the higher temperatures.

The figures in Table I show that this is the case, but the effect is rather hidden by the presence of erratic variation. In fact, the analysis of variance showed that the effect we are looking for, the weight versus temperature interaction, had a variance ratio of only 1.44, as compared with the weight versus temperature versus machine interaction, which is non-significant, representing a P of between 0.15 and 0.2. The effect can nevertheless be clearly shown by taking the ratio of the scores at the two higher temperatures to those at the lower two, at each value of weight separately. This gives the results shown in Table II; which show that the increase in error score

TABLE II

Handle load, lb. 	2	8	16	24	32	40
Score ratio 	1.10	1.09	1.11	1.15	1.14	1.19

brought about by an increase in room temperature from about 80° on the Effective Temperature scale to 90° Eff. is almost doubled by increasing the handle weight from 2 to 40 lb. It is interesting to compare this with Mackworth's results (Mackworth, 1950). Using the same apparatus he found that the effect of temperature was greater with 50 lb. loaded upon the control handle than it was with the heavy handle removed, and the apparatus controlled by a light knob operated by one hand only.

(d) *Measurements of Body Temperatures and Weight Loss.*

These measurements were taken as part of a longer-term programme whose aim is, if possible, to correlate change in performance with the physiological effects of heat. As was to be expected, and as we have found in previous experiments in the Cambridge hot-room, both body temperature and weight loss showed, when averaged for all the subjects, a close relationship to room temperature.

TABLE III

	Room Temperatures			
	85/75	90/80	95/85	100/90
Rise in body temperature, ° F. ..	0.45	0.55	0.9	1.2
Loss of weight, gm. 	350	500	800	1,100

The relationship between temperature and weight loss and individual performance scores is, however, slight. Product moment correlations of the physiological measures, with the extent to which each individual's score on any one day differed from his own average, gave figures of $r = 0.35$, with rise in rectal temperature, and $r = 0.15$ with weight loss. Of these, only the first is statistically significant, and that only just so. The relationship between the individual physiological measures and actual performance score, uncorrected in the above way, was still less. It is likely that, such as it is, this correlation is entirely due to the common factor of room temperature, and that there is no direct relationship between performance and either body temperature or rate of sweating. Although the figures from this experiment are far from sufficient in themselves to decide this point, such repeated negative findings suggest that the deterioration in performance is related to room temperature via some mechanism which does not involve a change in such physiological conditions as body temperature or rate of sweating.

VII

THE SUBJECTS' OPINIONS

Between 2 and 5 days after the experiment ended, ten of the twelve subjects were asked to complete a simple questionnaire form, asking for their opinions about the temperature and weight conditions. Comparing these opinions with the results obtained, it is interesting that:

- (a) eight subjects preferred some weight on the handle to none;
- (b) the largest weight which it was thought could be controlled without increase in error was between 24 and 32 lb., whereas, in fact, 24 lb. produces an increase in error;
- (c) eight subjects thought 90/80 was the highest room temperature they could tolerate without increase in error, the other two considered 95/85 was tolerable without decrease in accuracy. The majority vote in this case agrees with the numerical results very well.

VIII

CONCLUSION

The results of this experiment are in themselves insufficient to show the interaction of handle weight and room temperature to be statistically reliable, and thus they do not add any certainty to the inherent probability of such an effect. In providing a comparison between the effects of handle load and room temperature, however, the result of this experiment was satisfactory.

The most interesting item of information gleaned from the questionnaire replies is the confirmation that the addition of a small weight—about 10 lb.—to a control handle of this type, does lead to some improvement in accuracy of tracking.

IX

SUMMARY OF RESULTS

(1) Earlier findings, that performance of the tracking task provided by the pursuitmeter was adversely affected by high environmental temperatures, were confirmed.

(2) The magnitude of the effects of heat and of load are of the same order, and, in fact, the deterioration in accuracy between the two lowest and the two highest temperatures is equal to that between weight loads of 8 and 30 lb., i.e. an increase

of effective temperature from 80° to 90° is equivalent to adding 20 lb. load to the handle.

(3) The results indicate that the effect of loading the handle is greater at higher room temperatures, but this effect is not statistically reliable in this experiment.

(4) Rectal temperature and amount of weight lost in sweat show the expected relationship with room temperature, but the individual readings show no reliable correlation with performance.

X

REFERENCES

1. MACKWORTH, N. H. (1950). "Researches on the measurement of human performance." Medical Research Council, Special Report No. 268.
2. MACKWORTH, N. H. (1946). Effects of heat and high humidity on wireless telegraphy operators' hearing and recording Morse messages. *Brit. J. Indust. Med.*, **3**, 143.
3. HICK, W. E. (1946). "The effect of heavy loads on handwheel tracking." Unpublished report, RNP46/313, to the Royal Naval Personnel Research Committee of the Medical Research Council.

(Manuscript received 17th June, 1949.)

AMNESIA AND THE GENERIC IMAGE*

BY

O. L. ZANGWILL

(From the Institute of Experimental Psychology, Oxford)

It is pointed out that the fate of generic images has been studied little, if at all, in amnesias of organic origin. Observations are described in a case of alcoholic Korsakoff psychosis which appear to indicate that generic images are subject to Ribot's Law of Regression. The patient, a spinster of 57, with some talent for drawing, was asked to represent a number of objects, e.g. a woman dressed in the latest fashions, which might reasonably be regarded as "generic" in nature. These drawings suggested that the patient's generic images were, in some cases at least, appropriate to an earlier period of her personal history. In general, this ante-dating effect was in good correspondence with the scope of the retrograde amnesia. Some implications of these findings for the psychological theory of memory are briefly discussed.

I

INTRODUCTION

THE origin and nature of generic images were much discussed in the earlier systematic texts. "An image," wrote Stout, "is called generic when it possesses a distinct and salient centre or core corresponding to the common characters of a class, together with a vague and inconstant margin" (1896, I, p. 179). In particular, the possible role of the generic image in the evolution of the concept aroused lively controversy. This problem, which was admirably discussed by William James (1890, II, pp. 45-50), need not concern us here, and it is sufficient to recall Ward's comment that "The 'generality' of the generic image differs from the true universality of the concept in that the former is the passive and accidental result of reduplication, the latter the product of definite and active comparison" (1918, p. 199). As regards the origin of the generic image in individual experience, it has commonly been held to arise from the blending or fusion of a large number of particular images of identical or related content. As is well known, this view gained considerable prestige as a result of Galton's studies of composite portraiture (1907, pp. 229-33). It is highly doubtful, however, whether the parallel between the generic image and the composite photograph is as close as Galton supposed. In the first place, more recent studies (e.g. those of Katz, 1948) suggest that composite portraits are apt to acquire characteristics lacking in the typical generic image. And in the second place, it is now customary to regard recollection as displaying a more active character than is implied in Galton's treatment.† At the same time few would disagree with Galton's dictum that the character of our generic (or "cumulative") images depends, to a considerable extent at least, upon our nurture (1907, p. 132).

The changes undergone by generic images in the course of life would repay closer study. It is evident, for instance, that the generic images which we entertain in childhood differ appreciably in content from those which we entertain in adult life. Further, our general image of a given individual (or class of individuals) undoubtedly changes in content as we ourselves grow older. If, further, it is agreed that the images in question depend on some kind of cumulative disposition, it might be

* Based on a paper read to the Experimental Psychology Group on 20th July, 1947.

† "Wherever a single cause throws different groups of brain elements simultaneously into excitement, the result must be a blended memory."—Galton (1907, p. 229.) This was written in 1879.

expected that they would show regressive changes in cases of organic amnesia associated with brain injury or disease. The purpose of the present paper is to report some observations in a case of Korsakoff's psychosis which appear to confirm this expectation.

It has long been recognized that organic amnesia displays a retrograde (or retrogressive) tendency. This was formulated by Ribot under the title of the *law of regression* (1885, pp. 121-2). In cases of diffuse brain disease, recent events are the first to be forgotten and, as the disease advances, the amnesia comes to embrace events lying further and further back in the patient's psychological history. In severe cases of Korsakoff's psychosis, it is usual to find a retrograde amnesia whose scope embraces a very considerable proportion of the patient's life history. In the examination of such cases, however, it has been customary to restrict inquiry to the patient's memory for events in his personal life. To the knowledge of the present writer, no investigator has attempted to study the fate of generic images in patients who have sustained prolonged retrograde amnesias. It is felt, therefore, that the present observations, slight as they are, warrant being placed on record.

II

OBSERVATIONS IN A CASE OF KORSAKOFF'S PSYCHOSIS

The observations now to be described were made in connection with a more extensive investigation of amnesic syndromes undertaken in 1939-40 and interrupted by the war. The work was carried out at St. Ebba's Hospital, Epsom,* and at the Cambridge Psychological Laboratory. Some reference to the present case has already been made by the writer (Zangwill, 1941) in a paper on paramnesic phenomena in the Korsakoff state.

The patient was a spinster, aged 57. On admission (June, 1939) she was confused, disoriented and grossly amnesic. She made frequent mistakes of identity and confabulated freely. She was in poor general health, with peripheral neuritis in both legs. The mental condition remained stationary for some six months, after which time a certain improvement was manifested. Orientation gradually returned and confabulations became rarer and less spontaneous. At the time our observations were made the patient was approximately oriented in all spheres, but displayed a gross defect of recent memory and a long, though somewhat ill-defined retrograde amnesia. The scope of the latter embraced at least ten years of the patient's life. Insight was just beginning to return, but the patient entirely failed to appreciate the full gravity of her memory defect. She showed the usual dullness and lack of spontaneity of the Korsakoff case.

It was unfortunately not possible to obtain a full previous history in this case. It was ascertained, however, that the patient had had an ordinary elementary education and had then taken a full secretarial training course. She had worked for many years as a secretary with various firms, at one time holding a reasonably responsible clerical position in a bank. She had artistic and literary tastes, together with some slight talent for sketching. Intelligence as measured by tests indicated that the basic level was well above average, but there had been considerable deterioration of the standard organic type. By temperament, the patient was reserved and seclusive, but there were no indications of paranoid or obsessional features in the previous personality.

The diagnosis was *alcoholic Korsakoff psychosis*, with polyneuritis.

Analysis of Retrograde Amnesia.—In view of the fragmentary history it was impossible to reconstruct with any confidence the actual course of the patient's life during the years preceding the onset of the psychosis. It appeared plain, however, that remote memory (i.e. for events of childhood and early adult life) was intact,

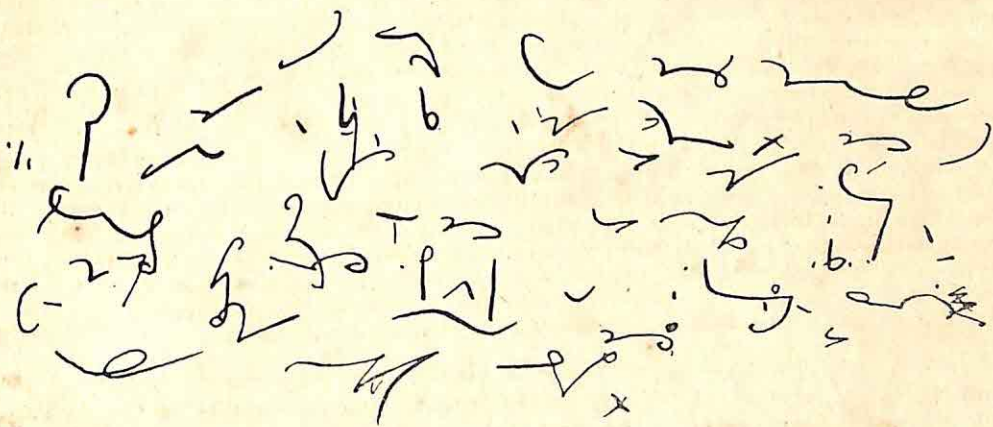
* I wish to express my sincere gratitude to Dr. L. H. Wootton, formerly Medical Superintendent, St. Ebba's Hospital, for his kindness in permitting me to study this patient.

although the chronology appeared somewhat confused. Outstanding events in the patient's personal life up to about the end of the 1914-18 war were accurately recalled, but there was marked uncertainty regarding her movements in the early post-war years. Amnesia appeared to be essentially complete for the fifteen years preceding the onset of the illness, but it was impossible to define its scope with complete conviction.

The prolonged retrograde amnesia, and the lack of full insight into it, gave rise to characteristic distortions of judgment. Thus the patient consistently underestimated her age by some 10 or 11 years despite the fact that she knew the actual year and could recall her year of birth. She was never fully convinced by arithmetical demonstration of her true age. She was also prone grossly to under-estimate the time that had elapsed since events had taken place which she was still able to recall. Thus she would regularly say that it was only 3 or 4 years since the Armistice. On the other hand, her knowledge of current events was not wholly at fault. Thus she believed that the present King was the "Duke of York" and she had some vague appreciation of the fact that the country was again at war with Germany.

It was of some interest to inquire whether the patient had retained her knowledge of *shorthand* and some dictation tests were accordingly given. One sample of her work is reproduced in Figure 1. The text of this passage reads as follows: "She called upon every women's organization throughout the country to devote itself to war work. Women, she added, were not being fully employed in the war effort. For instance, there were no women on the management committee of the twenty-seven childrens' camps set up under the evacuation scheme, though there are many women with the necessary managerial experience." An experienced stenographer to whom this sample was submitted reported as follows: "I should say that this was the work of a comparative beginner at shorthand, because many of the vowel sounds are put in which may later be left out in practice." Although it is possible that this feature indicates some regression to an earlier level of performance, it must be borne in mind that the patient had not been actively engaged in writing shorthand for some years. In any case, she had learnt shorthand as a girl and could readily remember doing so. The case is not, therefore, parallel to that reported by Stern (1938, p. 252), in which the patient "... could write shorthand, which he had learned during the forgotten period, with undiminished skill."

FIGURE 1



The patient's capacity for *sketching* appeared to have undergone little or no deterioration, but it was not possible to gain access to work which she had executed prior to her admission. The general level of her free-hand drawings was very much above the average.

Observations on Generic Images.—A number of simple tests were given with a view to sampling the patient's generic imagery. Some of the more important of these may be briefly described.

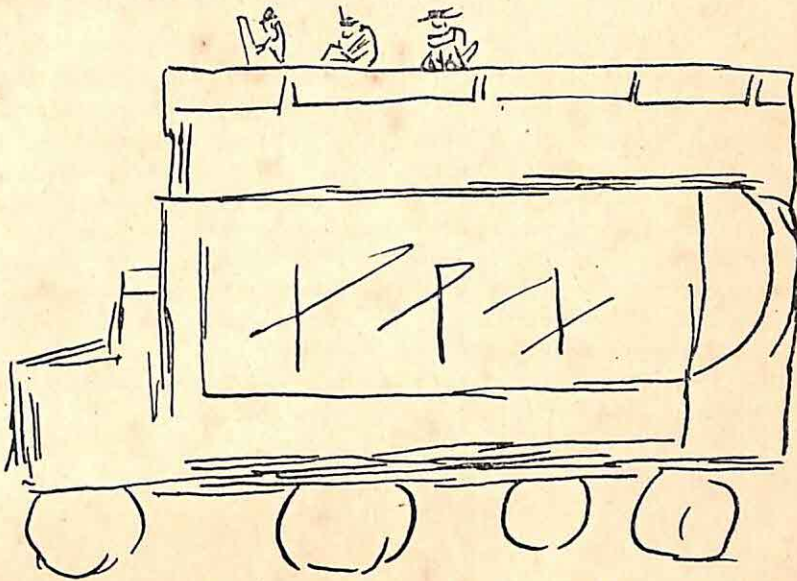
Test 1.—The patient was asked to think of a woman such as she might see any day in the street outside her home. She was then asked to draw a woman such as the one she had just imagined. She was particularly asked to represent the *latest fashion in dress* which she could remember. One of her drawings, which is typical of many obtained under these conditions, is shown in Figure 2. On this occasion, as on all others, the patient was entirely satisfied that *the style of dress represented in the picture corresponded in all essentials with the latest style she could remember and that this represented the "latest fashion," i.e. that of the year of her admission to hospital (1939).* The drawing shown in Figure 2 was submitted to a number of normal subjects of the same sex, and of approximately the same age, as the patient, and the style of dress depicted was on no occasion said to be more recent than 1930. Indeed, most judges dated it somewhere between 1917 and 1924.

FIGURE 2



Test 2.—The patient, who was a Londoner, was asked first to visualize, and then to draw, a bus of the type habitually seen in London during the year preceding her admission to hospital. It was found that she would invariably draw a double-decker bus *with an open top* and alleged that *all double-decker buses presented this appearance.* It will be borne in mind that open-top buses were becoming obsolete in London in the late 1920's and were finally withdrawn from general circulation in the early 1930's. One of the patient's drawings of a bus is shown in Figure 3.

FIGURE 3



Various other tests of the same kind were given with broadly similar results. In general, it appeared that the patient's recollection of objects which had been subject to progressive change in fashion or design had undergone a regression in accordance with the scope of her amnesia.

III

DISCUSSION

It has long been appreciated that the Korsakoff patient regularly interprets his *milieu* in terms of an environment familiar to him at an earlier stage of his life. Although, in the present case, orientation was largely restored by the time the patient came under our observation, there can be little doubt that her outlook remained markedly influenced by the scope of the amnesia. As we have said, it appeared to her to be no more than 3 or 4 years since the end of the Great War and she consistently (and without conscious intention) under-estimated her age by at least 10 years. In much the same way it would appear that her generic images of classes of objects or individuals were, in some cases at least, appropriate to an earlier period of her personal history. She appeared unable to call up generic images built up within the previous 10 years and presumably based on the integration of more recent experiences. It is probable, too, that memory images of individuals displayed the same regressive character as those of more general content. Although no decisive evidence could be obtained in the present case, observations on similar cases certainly point to this conclusion. In short, we may tentatively conclude that *memory images, in their generic or "cumulative" aspects, are subject to the law of regression.*

One may suggest that generic images in the normal individual are constructions based upon a large number of discrete past experiences. If we care, with Bartlett (1932), to employ the term *schema* to denote an active organization of related past experiences, we may say that the generic image is a function of the schema operating *en masse*. But it is plain that more recent experiences contribute predominantly to the content of the image. In the normal individual, indeed, recall of an earlier

and now out-dated fashion in dress or design is notoriously exacting. It is likely, therefore, that our patient, prior to the onset of her illness, would have found it extremely difficult to represent a woman in the fashions of 15 years before. One may suggest that, as an indirect consequence of the amnesia, there is a *release* of memory dispositions (or schemata) built up at an earlier period of life. Indeed, it has long been suggested that past experiences at the threshold of an amnesic gap in the patient's personal memory exercise a prepotent influence in determining the content of disorientation. There is a type of fixation upon an earlier period in the patient's life (Meggendorfer, 1928). Regression in the content of generic images has, therefore, both a negative and a positive aspect. The negative aspect is represented in the failure of recent memory; the positive aspect in the release of schemata normally superseded and held in check by the impact of current events. In consequence, the generic images evoked tend to be in keeping with an earlier era in the psychological history of the patient.*

* The finding that generic images are subject to retrograde amnesia may account for part, at least, of the bewilderment which Korsakoff cases typically display in their dealings with the environment. The confusion attendant upon a lack of correspondence between individual objects and the appropriate generic image may readily be imagined.

IV

REFERENCES

1. BARTLETT, F. C. (1932). *Remembering*. Cambridge.
2. GALTON, Sir Francis (1907). *Inquiries into Human Faculty*. Everyman's Edition. London.
3. JAMES, W. (1890). *Principles of Psychology*. London.
4. KATZ, D. (1948). Paper read to 12th International Congress of Psychology, Edinburgh. (Unpublished.)
5. MEGGENDORFER, F. (1928). "Intoxikationspsychosen" in Bumke, O. *Hdbh. d. Geisteskr.*, 7, 272-85.
6. RIBOT, Th. (1885). *Diseases of Memory*. London.
7. STERN, W. (1938). *General Psychology*. New York.
8. STOUT, G. F. (1896). *Analytic Psychology*. London.
9. WARD, J. (1918). *Psychological Principles*. Cambridge.
10. ZANGWILL, O. L. (1941). On a peculiarity of recognition in three cases of Korsakow's psychosis. *Brit. J. Psychol.*, 31, 230-48.

(Manuscript received 10th June, 1949.)

STUDIES IN SPACED AND MASSED LEARNING:

III. PAIRED-ASSOCIATE AND SERIAL LEARNING

BY

J. C. TSAO

(From the University of Hong Kong)

An experiment on paired-associate learning was carried out to test Hovland's prediction that spaced practice should be more advantageous than massed if the order of presenting the pairs of nonsense syllables were fixed.

In the results, spaced and massed practice showed little difference in learning efficiency. Hovland's prediction was not confirmed.

In a previous experiment of this series of studies it had been found that spaced practice was more efficient than massed in serial learning when syllables of "low-meaningfulness" value were used; but that there was little difference in efficiency between the two forms of distribution when "high-meaningfulness" syllables were used. The results of the two experiments naturally suggested that, as regards the relative efficiency of spaced and massed practice, paired-associate learning of low-meaningfulness syllables differed from serial learning of similar syllables, but matched serial learning of high-meaningfulness syllables. It seems justifiable to draw the conclusion that the relative efficiency of spaced and massed practice is different in paired-associate and serial learning, not because remote associations can be formed in the one case and not in the other, as Hovland (1939a, 1939b) assumed; but because the organization of the material and the learning processes are different in the two forms of learning.

I

INTRODUCTION

In two experiments, Hovland (1939a, 1939b) found that reminiscence was very marked, and distributed practice was superior to massed practice in serial learning; but that this was not so in learning paired associates which were presented in a constantly changing order. Hovland explained the results by assuming that fewer remote associations could be formed with paired associates and that consequently there was less inhibition, and thus reminiscence effects and the advantage of spacing, both of which depended on the dissipation of inhibition during interpolated intervals, were naturally less conspicuous in paired-associate than in serial learning. He predicted that, if the order of presentation of the successive pairs was not varied, reminiscence effects and, by implication, the advantage of spaced practice, would be intermediate between his results and those in serial learning.

McClelland (1942), and even Hovland (1939a) himself, suspected that the absence of reminiscence and the small advantage of spacing in Hovland's results might be due to the exposure time which was 2 seconds for each member of a pair. Neither Hovland nor McClelland found reminiscence at a 4-second rate of presentation. Hovland, however, thought that this result was unlikely, and McClelland did not subject his hypothesis to test.

Hovland's prediction concerned the order of presentation, fixed or varied; and McClelland suspected the effects of 4- or 2-second exposure time. The "prediction" and the "suspicion," put together, lead to an interesting problem: if paired associates were learned with a 2-second exposure time for each pair and a fixed order of presenting the successive pairs, would spaced practice be much more efficient than massed, just as in serial learning? In order to answer this question the present experiment was carried out.

II

METHOD

Material.—Two lists of eleven pairs of nonsense syllables each were used. The syllables were selected from Hull's (1933) list; all of them were of "meaningfulness value" from 0 to 5 on Hull's scale. To each list an additional pair was attached, which served as an example. The two lists are as follows:

<u>LIF</u>	<u>KUJ</u>	<u>LED</u>	<u>HIJ</u>
	<u>KUJ</u>	<u>LED</u>	
WIB	FOV	BIW	ZUD
HAJ	YUT	VEF	ZOJ
JID	VAK	PIJ	TEV
WOJ	FEP	MUJ	VIP
ZIH	VOF	WOF	ZIK
VAW	KOY	KEF	HUJ
ZUX	MIV	YOJ	VUK
FUP	BEJ	FAP	YUF
VEP	BUW	POB	MAF
ZOY	TIV	TUJ	VAB
KEJ	ZIB	KIH	MEV
	FOV	BIW	
	HAJ	VEF	
	JID	PIJ	
	FEP	VIP	
	VOF	WOF	
	VAW	HUJ	
	MIV	VUK	
	FUP	FAP	
	BUW	MAF	
	TIV	TUJ	
	KEJ	MEV	

Subjects.—The subjects were twenty-four men, twenty of whom were R.A.F. personnel and four R.N. staff. None of them had any previous experience of learning nonsense syllables or knew the purpose of the study before or during the experiment.

Procedure.—The general procedure was similar to that described in the first article of this series (Tsao, 1948a). The same memory apparatus was used. The "example" pair of the first list was first exposed and explained to the subject; the eleven pairs followed the example. Each pair was exposed for 2 seconds with an interval of $\frac{1}{2}$ second between each successive pair. The single syllables followed the eleventh pair, and the subject was asked, on seeing a single syllable, to pronounce the other one which had been paired with it. Ten trials were given of each list. The order of pairs and single syllables was the same throughout the experiment as shown in the lists given above. In spaced practice, a 1-minute interval was interpolated between successive trials, during which the experimenter talked with the subject on current events or other general topics. In massed learning the readings were continuous. In order to balance practice effects and the influence of shifting from spaced to massed practice or vice versa, the learning of the two lists was arranged in such a way that each was learned, by different subjects, for an equal number of times in every possible situation in relation to each other and to the distribution of practice.

III

RESULTS

TABLE I

SCORES (MEAN NUMBER OF CORRECT RESPONSES) IN PAIRED-ASSOCIATE LEARNING BY SPACED AND MASSED PRACTICE

Practice	Trials									
	1	2	3	4	5	6	7	8	9	10
Spaced ..	0.46	0.75	0.96	1.46	1.46	1.67	2.21	2.17	2.54	2.92
Massed ..	0.42	0.83	1.00	1.21	1.25	1.38	1.75	2.17	2.25	2.71

The learning score of each subject in each trial is the number of correct responses. Table I contains the average scores of each trial in spaced and massed practice. Since there were twenty-four subjects and each of them learned once by spaced practice and once by massed practice, each average score is based on twenty-four individual scores. The results are also presented in a graphic form in Figure 1.

FIGURE 1

LEARNING CURVES OF SPACED AND MASSED PRACTICE IN PAIRED-ASSOCIATE LEARNING

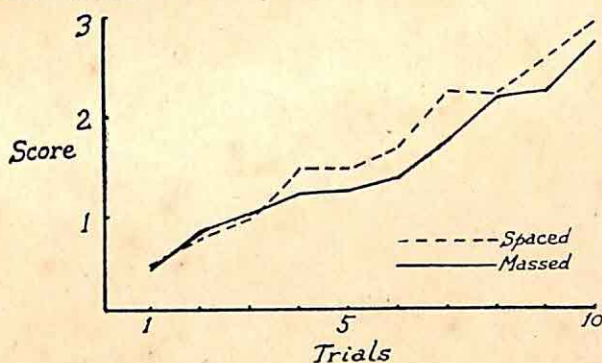


TABLE II

SCORES OF, AND DIFFERENCES BETWEEN, SPACED AND MASSED PRACTICE IN PAIRED-ASSOCIATE LEARNING

	Spaced					Massed			
10th Trial:									
M.	2.92				2.71
Dif.		0.21			
C.R.		0.44			
9-10 Trials:									
M.	5.46				4.96
Dif.		0.50			
C.R.		0.55			
1-10 Trials:									
M.	16.58				14.96
Dif.		1.62			
C.R.		0.53			

From the data in Table I and the curves in Figure 1, it can be seen that the efficiency of massed practice is generally on a lower level than that of the spaced learning, but the discrepancy between these two forms of practice is small. The differences, and the corresponding C.R.'s, between the spaced and the massed practice in three measures, viz. mean scores of the 10th, 9-10th and all trials, are given in Table II. None of the differences is statistically significant. The results show, therefore, that in paired-associate learning spaced and massed practice made very little or no difference in learning efficiency, or more concretely, that 1-minute intervals interpolated between successive trials did not greatly affect learning efficiency in paired-associate learning.

IV

DISCUSSION

The results of the present experiment are quite different from those of serial learning of low-meaningfulness syllables in another experiment of this series (Tsao, 1948b), which showed that spaced practice was much more efficient than massed. This may be an indication that there is some difference between paired-associate and serial learning, which caused the difference in efficiency of distributions of practice in the two types of learning.

The technique used in the present experiment was different from Hovland's (1939a) in the following aspects: (1) In the present experiment the order of pairs in presentation was fixed while Hovland constantly changed it. (2) The exposure time for each pair was 2 seconds in this experiment while in Hovland's it was 4 seconds. (3) The two members of each pair were exposed simultaneously inside the same window in the present experiment, while in Hovland's the two members were shown successively and alternatively in two windows on his apparatus. Hovland predicted that a constant order of presentation of pairs would induce remote associations and render spaced practice advantageous because of the dissipation of inhibition during interpolated intervals. Both Hovland and McClelland suspected the possibility that, with a 2-second exposure time for each pair, reminiscence effects and, by implication, the advantage of spaced practice, would be found. Hovland (1939a) thought that, in his experiment, "the continuous rotation of the order of the pairs and the alternative exposure of the syllables in different windows of the apparatus is less monotonous to the subject, so that decremental effects are less in paired-associate lists than in serial learning where the syllables are presented in the same order, in the same window and at the same rate." In fact, the results of the present study do not confirm any of Hovland's or McClelland's suppositions.

Hovland assumed that the difference between paired-associate and serial learning lay in the formation of remote associations. Remote associations, according to Hovland, were usually formed abundantly in serial learning, but not in paired-associate learning in which the order of presentation of the pairs varied constantly. Time intervals, during which inhibitions due to remote associations dissipated, were, therefore, more beneficial in serial than in paired-associate learning. In paired-associate learning with a fixed order of presentation of the pairs, Hovland predicted remote associations would be formed. This part of Hovland's prediction was right; it was noticed in the present experiment that the subject responded very frequently to a single syllable with a syllable other than its partner in the list, sometimes even with a syllable in the other list. Hovland's prediction broke down, however, in the

other part; spaced practice was not much superior to massed in paired-associate learning with a fixed order of presentation of the pairs, although plenty of remote associations had been formed, as is shown by the results of the present experiment.

With the failure to confirm Hovland's prediction, his hypothesis of inhibition apparently cannot be used to explain the fact that spaced practice was much more efficient than massed in serial learning of low-meaningfulness syllables, but was not so in learning paired associates of similar material as shown in the present and the other experiment (Tsao, 1948b). The differential effects of distribution of practice may, instead, be caused by some differences both in the organization of material and in the learning processes of the two forms of rote learning.

In serial learning a list is just a chain of nonsense syllables. Its items, at the beginning, are homogeneous to the subject. This, as noticed by Restorff (1933), makes serial learning a difficult task. Only as the learning proceeds, grouping and other devices, as observed by Müller and Schumann (1894) and later investigators, are developed by the subject, which make the syllables more accessible and, furthermore, facilitate the learning. Paired associates, on the other hand, provide some organization in the material. From the very beginning the syllables are given in pairs which may be taken as some sort of "sub-wholes." Since each syllable in paired associates is "matched" by another one and associated with it, it is thus given an "artificial meaning." So the material in paired-associate learning may be regarded as better organized and even a little more "meaningful" than the material in serial learning. Meaningful materials are usually learned about equally well with spaced and massed practice, as has been shown in a former experiment of this series (Tsao, 1948b).

The above refers to the organization of the material; there are also some differences in the learning processes in paired-associate and serial learning. Both in paired-associate and serial learning a syllable has to be memorized as a single syllable. Also, in serial learning a syllable has to be memorized in its serial position, that is, before and after certain syllables, especially after a certain syllable. In paired-associate learning, a syllable has to be remembered with a certain partner which forms a pair with it and serves as the cue for its recall. In serial learning, with the anticipation method, the subject must go back to the preceding syllable in order to give a correct response. He has to remember the preceding one as well as the very syllable concerned, before he can anticipate it correctly. Moreover, a syllable is remembered not only as a syllable itself, but also as a cue for recalling the following one which is still to come. The syllables are exposed individually, but the subject sees only one syllable at a time, and he must relate it to both the preceding one which is no longer present and the succeeding one which is not yet in view. Serial learning is, therefore, a rather complex process. Paired-associate learning is different. A pair of syllables is shown at a time. In testing the memory, one of them is given; what the subject has to do is only to recall the other one. Since in learning both syllables are exposed simultaneously, the subject does not have to associate one of them with something not in view. In comparison with those of serial learning, the processes in paired-associate learning are much simpler. A more complicated learning process is, of course, a more difficult one, and thus serial learning and paired-associate learning differ in difficulty. That the advantage of distribution of practice is more prominent in difficult tasks than in easy tasks has been proved by Cook (1928), Borovski and Kobozeva (1935) and many others.

This study was carried out in the Psychological Laboratory, Cambridge. I wish to express my gratitude to Mr. G. C. Grindley for help and advice.

V

REFERENCES

1. BOROVSKI, V. M., and KOBOZEVA, P. S. (1935). Relation between distribution of practice and difficulty of the problem. *Instinkt. Naviki.*, **1**, 68-76.
2. COOK, S. A. (1928). The effect of various temporal arrangement of practice on the mastery of an animal maze of moderate difficulty. *Arch. Psychol.*, **15**, No. 98.
3. HOVLAND, C. I. (1939a). Experimental studies in rote learning theory: IV. Comparison of reminiscence in serial and paired-associate learning. *J. Exp. Psychol.*, **24**, 466-484.
4. HOVLAND, C. I. (1939b). Experimental studies in rote learning theory: V. Comparison of distribution of practice in serial and paired-associate learning. *J. Exp. Psychol.*, **25**, 622-633.
5. HULL, C. L. (1933). The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, **45**, 730-734.
6. MCCLELLAND, D. C. (1942). Studies in serial verbal discrimination learning: I. Reminiscence with two speeds of pair presentation. *J. Exp. Psychol.*, **31**, 44-56.
7. MÜLLER, G. E., and SCHUMANN, F. (1894). Experimentelle Beiträge zur Untersuchung des Gedächtnisses. *Zeit. Psychol. Ergbd.*, **1**.
8. RESTORFF, H. VON (1933). Über die Wirkung von Bereichsbildung im Spurenfeld. *Psychol. Forsch.*, **18**, 299-342.
9. TSAO, J. C. (1948a). Studies in spaced and massed learning: I. Time period and amount of practice. *Quart. J. Exp. Psychol.*, **1**, 29-36.
10. TSAO, J. C. (1948b). Studies in spaced and massed learning: II. Meaningfulness of material and distribution of practice. *Quart. J. Exp. Psychol.*, **1**, 79-84.

(Manuscript received 20th August, 1948.)

THE EFFECTS OF REPEATEDLY RETESTING THE SAME GROUP ON THE SAME INTELLIGENCE TEST:

II. HIGH GRADE MENTAL DEFECTIVES

BY

A. W. HEIM and J. G. WALLACE

(From the Medical Research Council Applied Psychology Unit, Cambridge)

Twelve mentally-defective schoolboys between 14 and 16 years of age took an intelligence test (AH 4) once a week, ten times. They took also Progressive Matrices once, after they had finished the AH 4 testing. Subjects of low intelligence were chosen in order to throw some light on the findings of a previous experiment on repeated retesting, in which the subjects were intelligent adults. The main aims of the schoolboy enquiry were to ascertain how far the flattening of the progress curves observed in the earlier enquiry was due to the artificial "ceiling" imposed and how far the part played by variation within the individuals varies with the level of intelligence of the group.

The mentally-defective subjects showed a gradual improvement in test performance with little sign of flattening towards the end; their individual progress curves were far less smooth than had been those of the more intelligent subjects. Other findings included relatively high proficiency on the diagrammatic part of the test, as opposed to the verbal and numerical, and some transfer of training from AH 4 to Progressive Matrices.

I

INTRODUCTION

In a previous report (Heim and Wallace, 1949) the results of repeatedly retesting a small group of intelligent adults once a week, for some weeks, were described. It was found that the typical individual progress curve showed a steep upward rise for the first four or five testings, followed by a levelling out towards the end. The order of the subjects remained strikingly constant, but the correlations between the two parts of the test tended to fall with successive testings, suggesting that the part played by "bias" increases with repeated testings. There was some evidence that Part I (verbal and numerical) presented less difficulty to the group than did Part II (diagrammatic).

These findings raised questions of theoretical and practical interest. Does intelligence test performance, like other skills, always improve with practice? Does rate of improvement always become negligible after the first few testings? Or are these findings dependent upon a particular relation between group and test used?—i.e. a pre-selected group whose intelligence level is higher than that of the population for whom the test was originally designed. Does the "learning" extend to other somewhat dissimilar tests? Can subjects within the group always be relied on to maintain their relative position or is this, again, dependent on the relationship between group and test used?

The solution of these and allied problems requires a series of experiments with different groups and different tests. Mental defectives were chosen in this instance, since it was felt that they would have the greatest scope for improvement: there would be little danger of their reaching the end of the test and gaining nearly full marks, even after many testings.

II

SUBJECTS AND TESTS

The subjects were twelve high-grade mentally-defective boys, ranging in age from 14.6 to 15.10, and drawn from Littleton House School, Cambridge. Only boys who were competent readers were accepted as subjects, in view of the largely verbal bias of the first part of Test AH 4. It was clear that any Part I scores gained by boys who have difficulty in reading simple words would be worthless.

The Binet I.Q.'s of the boys, established before they were admitted to the school, ranged from 55 to 76.5. The period between the Binet testing and the AH 4 testing varied from one to five years.

AH 4 testing was carried out weekly, for eleven consecutive weeks, at the end of which period the school term came to an end. In this way eight of the boys were tested eleven times and all twelve of them were tested ten times. Thus in four cases only there was on some one occasion a gap of two weeks between testings. The results discussed in this report are concerned, unless otherwise stated, with the ten tests taken by the whole group of twelve.

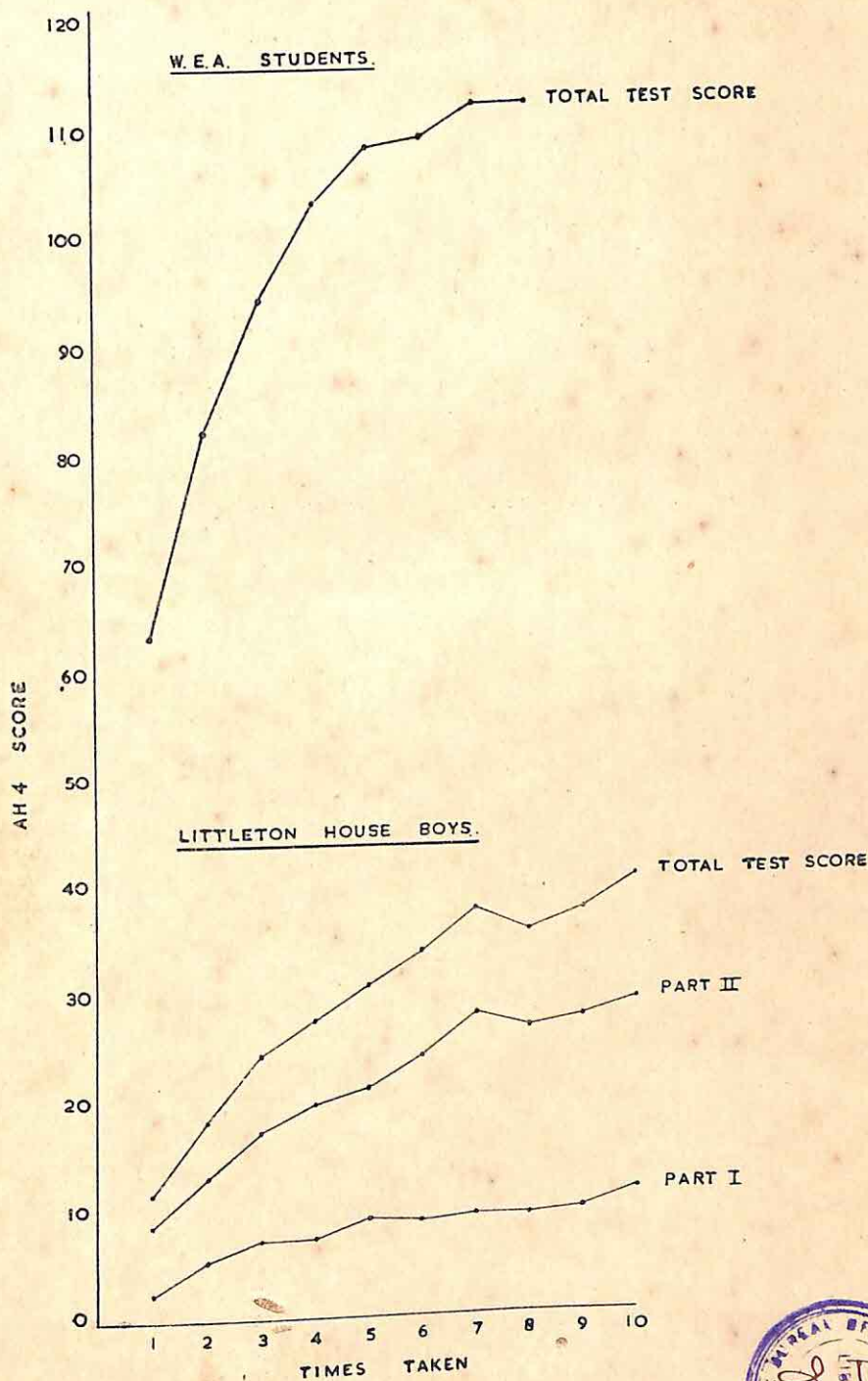
Every care was taken to ensure that the subjects understood what was required of them: the usual simple examples were given of the type of question met with in both parts of the test and the boys worked through these in their own time, each week, before embarking on the test proper, each part of which lasts 10 minutes.

The concentration of the mental defectives during testing was, of course, markedly inferior to that of the W.E.A. students who had taken part in the first enquiry. However, the boys appeared less subject to boredom in the course of testings—perhaps because they, unlike their predecessors, were always meeting new questions as they got further in the test throughout the ten weeks, perhaps because boredom with familiar things is a characteristic of high rather than low intelligence. The boys were interested and forthcoming; they were, for instance, very eager to see the stop-watch at the end of each testing and, if possible, to hold it. They were quite keen to hear whether they were improving as a group, and when they were told that they had done the test for the last time, they looked downcast and asked the tester when she would be coming back. The eleventh testing took place on 1st April, 1947. On that occasion several of the boys gleefully made mistakes over the Examples, in order to have the pleasure of making an "April Fool" of the tester.

At the end of April, after their school holiday, the twelve boys were given the Progressive Matrices Test. Two unusual points of procedure should be mentioned in connection with the latter testing. (a) Six examples of "Matrix" type devised by the tester were provided to be worked through, before the test proper was begun. The first and fifth of these had already been completed, but the boys were urged to study them; the remaining four examples had to be solved by the boys themselves. (b) A time limit of 30 minutes was imposed for the test proper. Five of the boys finished the test within this time and all but one of the others had reached set D (the fourth of the five sets of twelve questions).

In addition to the Binet I.Q.'s and the Progressive Matrices scores, the writers had access to the boys' school reports, which include remarks on performance in school subjects (including manual ones), speech, physique, behaviour and family history. The Headmaster kindly added his comments on the appropriateness, in his view, of the Binet I.Q. with which the boy had entered the school, and he also arranged the subjects in rank order according to his conception of their "innate all-round ability." The comparison of these criteria one with another yielded little of direct relevance to this particular enquiry. In order to save space, therefore, discussion of these results has been omitted from this paper.

FIGURE 1
PROGRESS CURVES OF MEAN SCORES



III RESULTS

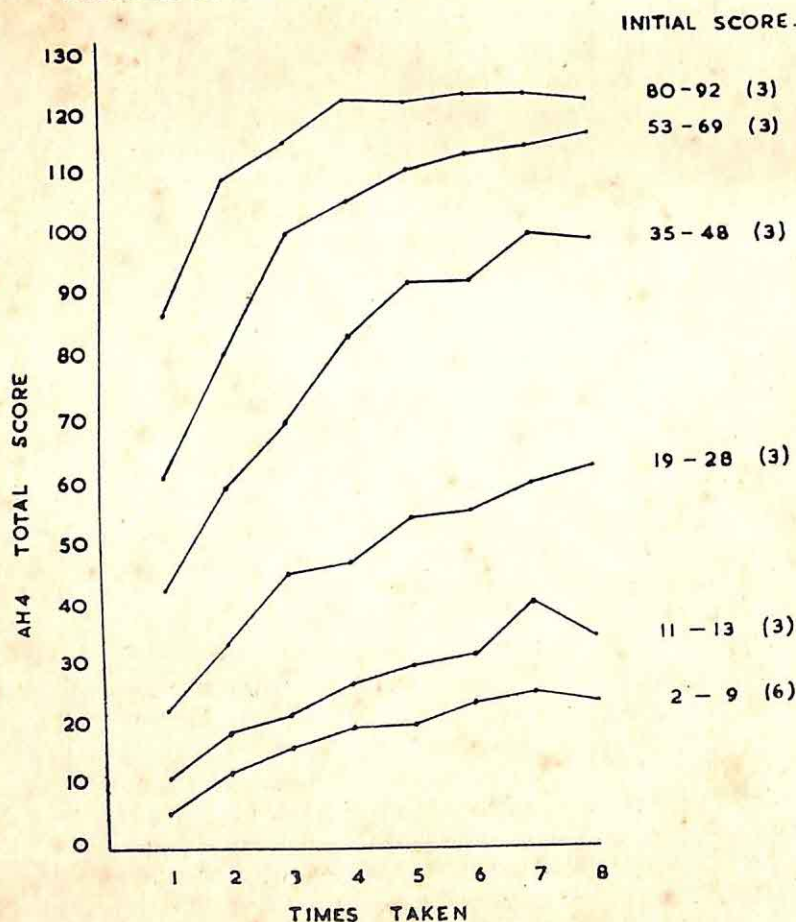
A. The General Trend of Scores.

Figure 1 shows the curves for the mean scores of the group of twelve subjects, for the ten testings. Separate curves are shown for Part I, Part II and total test scores. The total test score curve for the W.E.A. group is shown above.

It may be seen that the total score curves of the Littleton House boys and of the W.E.A. students differ considerably from each other. The W.E.A. curve rises steeply at first and begins levelling out somewhere around the fifth testing. The mentally-defective group, however, shows a fairly steady rise—except for trial 7 which yields a small peak.

This would suggest that the flattening out of the W.E.A. curve is a property of the relation between the test (intended for a cross-section of the population) and the group (a selected, intelligent group), rather than of the learning process as such. Given this relatively easy task, the W.E.A. group virtually finished all its "learning" after some five trials. The schoolboys' curve suggests that the duller subjects learn not only more slowly, but also over a considerably longer period, given the opportunity.

FIGURE 2
PROGRESS CURVES OVER 8 TRIALS AT DIFFERENT LEVELS



That the difference in the trends is a difference in degree rather than in kind is suggested by Figure 2 (for which idea we are indebted to Mrs. N. Harris). In this figure the results of the subjects from both groups over eight trials are included: each progress curve shows the mean for three subjects—adjacent on their first score—except for the bottom progress curve which is the mean of the *six* lowest-scoring subjects.

This figure illustrates clearly the change in gradient of the progress curve as the intelligence of the subjects decreases. It is observable that the "steps" taken by the brighter groups grow progressively smaller with each testing, whereas this is not true of the duller groups.

Inspection of the schoolboys' progress curves for the two separate parts of the test (Figure 1) throws some light on what is happening. The Part II scores are initially, and throughout subsequent testings, much higher than those for Part I; moreover, the improvement on Part II is far more marked than that on Part I. Linear regressions were fitted to the curves of Part I, Part II, and also of Part II minus Part I, in order to ascertain the improvement on Part II relative to Part I: these showed that the improvement is significantly greater for the second Part.

Thus it is clear that Part I contributes little to the improvement shown throughout the Total score curve: the shape of the Total score curve is very largely determined by that of the Part II curve. Evidently Part I is almost outside the intellectual range of the mental defectives. It appears unlikely that they would ever make a fair showing on Part I, even if they continued to take it for months, or if they tried taking it with unlimited time. On the other hand, their Part II scores suggest that they might well continue to improve if they went on taking Part II.

The initially higher scores and the greater progress made in Part II lends itself to various interpretations. It may be that unintelligent subjects tend to find diagrams intrinsically easier to deal with than words or figures. A more direct approach is possible with the diagrams: they represent nothing beyond themselves. But words and figures are one degree removed from what they represent; they are symbols to be read and translated before they can be used. It may be that at low levels of intelligence these processes present a real difficulty. Beyond a certain intelligence level this distinction evidently plays a far smaller part: intelligent subjects vary as to which "bias" they prefer.

A possible explanation of the difference in *slope* between the progress curves of the two Parts may lie in the fact that the subjects have had practice in school for some years, in dealing with words and figures, whilst they have had little or none with diagrams of the kind used in the test. They may therefore have made most of the progress of which they are capable in the first field, whereas they have considerably more scope—and, perhaps, interest—in the second.

That the difference between the Part I and Part II progress curve is not a simple matter, is evident from a comparison of the number of questions attempted with the ratio of number correct to number attempted, on the two Parts.

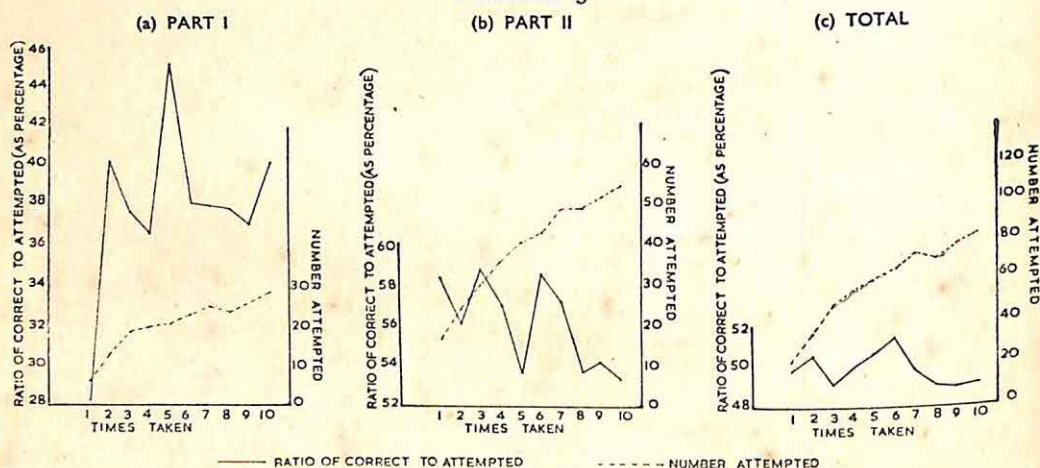
Figures 3 (a), (b) and (c) show these two curves for the two separate Parts of the test and also for the Total test score. It may be seen in Figures 3 (a) and (b) that, whilst the number of questions attempted increases to a greater extent and more rapidly for Part II than for Part I, the ratio of correct answers to number of questions attempted tends to decrease in Part II and, if anything, to increase in Part I—in so far as such serrated curves may be said to show a tendency.

This may be due to the fact that the questions in the test become progressively harder: in Part II the quicker boys were reaching quite difficult questions during

the last few testings, whereas none of them got very far in Part I. Furthermore, the Part II ratio was startlingly higher to begin with than the Part I ratio. The difference in trend between the two ratios might, of course, be due, to some extent, to the greater practice which the boys got on Part I, owing to their always attempting the *same* questions week after week.

On the other hand, for the W.E.A. subjects, the curve showing the number of questions attempted was similar in shape to their test score progress curve. But the ratio curves for the two Parts were diametrically opposed; in Part I the ratio sank steadily, and in Part II the ratio (*lower*, initially, than that for Part I) rose steadily—though it was never as high as the ratio for Part I. Thus the brighter subjects (with their somewhat verbally biased occupations) evidently found Part I absolutely, as well as relatively, easier than the diagrammatic second Part.

FIGURE 3



B. Individual Variation.

The scores on each testing were compared with the scores on every other testing and also with the "pool" of the scores. In correlating each set of scores with the pool, the set in question was included in the pool. The product moment method of correlation was used in this investigation as in the earlier one. It was chosen in preference to the method of rank correlation in order that occasional striking differences between particular individuals should not be masked and that a measure of the significance of differences would be available.

Using the scores of the whole Littleton House group, the test yielded very high consistency. The correlations between the separate testings and the "pool" ranged from 0.879 (second with pool) to 0.989 (fifth with pool). The correlations between scores on the different testings with one another were scarcely less close: with one exception (the second with the seventh, 0.709) they were all over 0.80, the majority exceeding 0.90.

In addition to the above comparisons between Total test scores, scores on the separate Parts of the test for each testing were correlated with their pool. These correlations were found to be very similar to those obtained with Total test score: they were all very high (over 0.785). No significant difference was found between the two Parts, and no trend was observable with repeated testings. Thus, all the correlations *between the testings* are highly significant, none of the differences *between the correlation coefficients* being significant.

Figure 4 (a) shows the individual progress curves of the twelve boys, numbered for purposes of reference. Inspection of these curves reveals that they fall fairly clearly into two separate groups, Nos. 1-4 in the first group and Nos. 5-12 in the second; and that the progress curves of the second group in particular cross and re-cross with great frequency. It was thought interesting, therefore, to examine the consistency of the mental defectives within these two groups. The correlation coefficients showed a very changed pattern; the range for the total group of twelve was $r = 0.709$ to 0.989 (a difference of 0.28), whereas that for the lowest eight was $r = -0.008$ to 0.917 (a difference of 0.925); only twenty out of the fifty-five correlation coefficients on the smaller group were significant; and a strong trend towards greater consistency was shown *with later testings*.

FIGURE 4 (a)
INDIVIDUAL PROGRESS CURVES
LITTLETON HOUSE GROUP

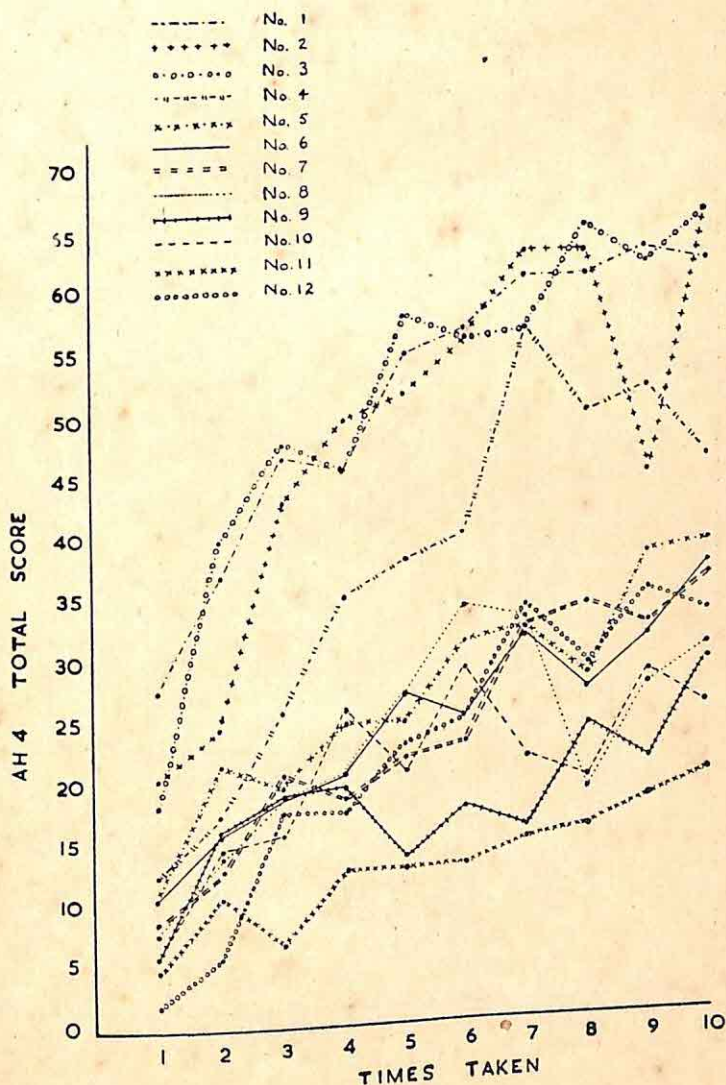
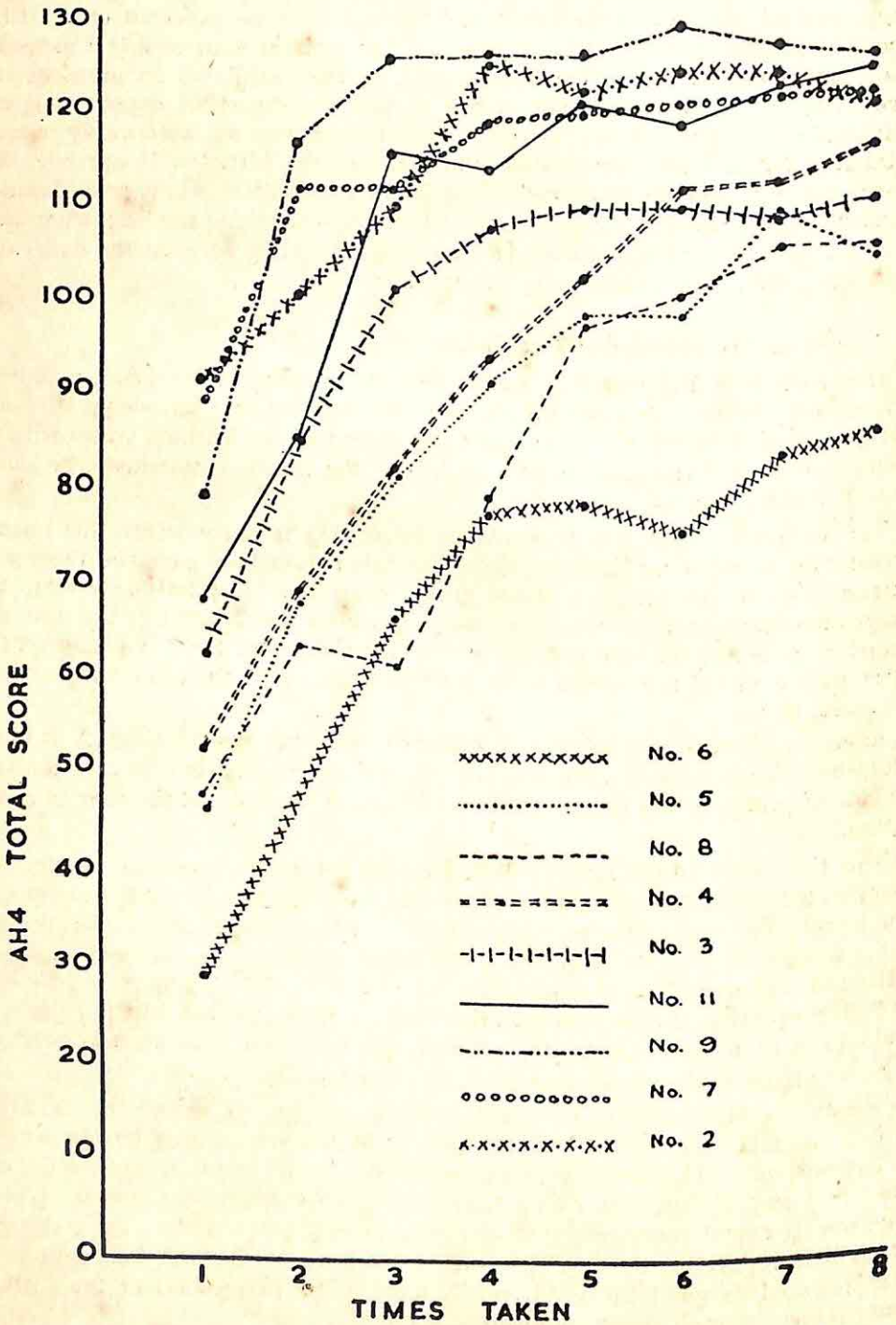


FIGURE 4 (b)
INDIVIDUAL PROGRESS CURVES
W.E.A. GROUP



An Analysis of Variance showed that significant differences in ability obtained between the subjects on both Parts of the test. It is clear, however, that the consistency of the total group is due largely to the fact that the first four maintained their position relative to the other eight throughout the experiment.

The relative inconsistency of the duller eight boys as compared with the total group, and as compared with the W.E.A. group, is to be expected on statistical as well as psychological grounds. Their range of score at each testing is extremely small; differences of one or two marks, such as they exhibited on initial testing, have little or no meaning; and chance played some part in determining each individual's score (see Section III (D)). Mental defectives are notoriously prone to instability and lack of concentration and many of the Littleton House boys were pre-selected for behavioural as well as intellectual difficulties. It is possible that a sufficiently low-grade test would reverse the finding—bunching the brighter subjects so that their inter-test consistency falls and discriminating between the duller ones more adequately and, hence, more consistently.

C. *"Learning" in Intelligence Test Taking.*

It is clear from Figure 1 that a good deal of "learning" takes place, as a result of repeatedly taking the same intelligence test, without any knowledge of results on the part of the subjects. The speed and extent of the learning varies with the intelligence level of the subjects and the bias of the test, but it seems to be always present in some degree.

The writers thought that it would be interesting to see whether this learning carried over to other intelligence tests. For this reason they gave the Progressive Matrices Test to the Littleton House group, when the AH 4 testing was finished. Progressive Matrices was chosen because it has been widely used in the past with groups of varied intelligence and because it was clear from the AH 4 findings that a test with a visual bias would yield more fruitful results than one with a verbal or numerical bias.

As noted above, examples were provided for the boys to work through at leisure before embarking on the test proper. The examples were easy, but they did illustrate the kind of principle that the subjects would meet in the test and the form of answer required.

The time limit of 30 minutes was imposed for purely practical reasons: the examples took some 15 minutes to complete and it was felt that longer than 45 minutes would probably overtax the boys' attention. In fact, the results showed that this timing was legitimate: the majority of the subjects reached the last section but one of the test and five of these finished the whole test in the 30 minutes. In view of the high proportion of errors made in the last two sections it was felt that the scores would have been substantially the same if unlimited time had been allowed.

None of the subjects had ever seen this test before.

The twelve Matrix scores were as follows: 41, 37, 36, 34, 30, 30, 25, 24, 21, 17, 16, 8. The following table shows two sets of Matrix norms used by the Army in the 1939-45 war. They are based on a sample of 5,000 representative of the war-time Army in terms of age, educational standard and non-verbal intelligence (Col. B. Ungerson, personal communication). The sample is not strictly comparable with the civilian population through the absence of those in Reserved Occupations and those rejected as mentally or physically unfit. The frequencies of the Littleton House scores for each grade are given in columns 4 and 6.

%	Grade	20-minute norms		45-minute norms	
		Matrix marks	Number of Littleton House subjects	Matrix marks	Number of Littleton House subjects
10	A	46+	0	54+	0
20	B	40-45	1	48-53	0
20	C+	35-39	2	44-47	0
20	C-	29-34	3	39-43	1
20	D	20-28	3	29-38	5
10	E	-19	3	-28	6

Thus the general standard is extraordinarily high for a group of subjects whose Binet I.Q.'s are all under 80. It may be seen that only half of the boys come within the bottom 10 per cent. of the Army population on the 45-minute norms, and only one quarter on the 20-minute norms.

In view of the difference in form between Progressive Matrices and even the visual-biased Part II of AH 4, this suggests some transfer of training. The corresponding grades on AH 4, on the first testing, were D, 2, and E, 10. On a bigger, more representative group, correlations of 0.66, 0.68 and 0.69 (Parts I, II and Total respectively) have been found between AH 4 and Progressive Matrices. It looked, therefore, as though the comparatively high Matrix scores were due largely to the "training" effect produced by repeatedly taking a different test, without knowledge of results.

There was a possibility, however, that these "high" scores were due to the presentation of Matrix-type examples to be correctly completed by the boys before they embarked on the Matrix Test proper. (See Wallace, 1949, Section II, 4, for influence of Examples on test scores.) A control group of Littleton House boys, comparable with respect to age and reading competence, were therefore tested the succeeding term on Progressive Matrices under exactly the same conditions as were the first group—save that the second group had never seen AH 4.

The scores of this group were as follows 40, 32, 27, 27, 23, 14, 14, 8, 7, 6, 6. The mean scores for the two groups were 26.6 for the test group and 18.5 for the control group. This is clearly a substantial difference, being a drop of over 25 per cent. It is not, however, statistically significant, $t = 1.8$ —instead of the requisite 2.08. These results are not, therefore, wholly conclusive: they suggest that the high Matrix scores of the first group of boys were due partly to the practice effect of repeated AH 4 testings and partly to the inclusion of the examples with the Matrix Test.

The effect of providing examples for the subjects to work through, and of test sophistication, can scarcely be overstressed in view of the increasing employment of psychological tests for practical purposes, and the practice of many psychologists of treating I.Q. as, within narrow limits, a reliable, immutable measure.

D. Differences between Groups of High and o. Low Intelligence.

The most striking difference between the W.E.A. students and the mentally-defective schoolboys was the difference in consistency of individuals. Figures 4 (a) and (b) show the individual progress curves for the two groups.

It may be seen that the high-grade subjects, with very few exceptions, yielded smooth curves which rarely crossed one another. The progress curves of the duller

subjects, however, oscillate wildly and repeatedly cross and re-cross one another. The erratic course of the individual boys is completely obscured in the progress curve of their means (see Figure 1). Most of the individual curves of the W.E.A. students strongly resemble their mean curve, whereas only one of the schoolboys (No. 1) produced a curve at all similar to the Littleton House mean curve. Discussions with the Headmaster, and with the boys after testing, threw no light on the violent changes in score from one week to another. No explanation was found, for instance, for 2's sudden drop at his ninth testing or for 4's impressive rise between sixth and seventh testing.

Inspection of the curves showing the ratio between correct answers and questions attempted showed the same difference between groups. The W.E.A. curves were smooth, they tended to fall for Part I and to rise for Part II. The Littleton House curves are so jagged that little can be deduced from them, but in so far as they exhibit any trends, these go in the opposite direction from those of the W.E.A. group.

The relationship between Parts I and II of AH 4 pursues the same oscillating course for the mental defectives—and the same relatively smooth one for the intelligent subjects, see Figures 5 (a) and (b). Whilst the smoothness of the W.E.A. curve is broken only twice (at the second and eighth testing), the Littleton House "curve" proceeds in erratic, irresponsible manner, to finish almost exactly where it began.

The general trend, however, is downward, as it is with the W.E.A. group: both the peaks and the dips, looked at separately, become progressively lower. In the W.E.A. report it was suggested that this decrease in relationship between Parts I and II might be due to the narrowing range of scores of the group, the standard deviation having been found to fall from 19.54 the first time, to 12.13 the last time. These new data suggest that this greater selectivity may not have been the main cause, since the standard deviation of the mentally-defective subjects tended to rise throughout the testings (first time 7.33, tenth time 15.24), no artificial ceiling being imposed in this instance.

It would appear then that the effect of individual differences in bias tends to increase with retesting, in any case when the test used fully extends the subjects. In view of the erratic nature of the curve in Figure 5 (a), however, further data on this point should be acquired, preferably with an unselected group of subjects.

The differences between the mean curves of Part I and Part II scores given in Figure 1, shows how important a rôle bias plays, whether the subjects are test-sophisticated or not, when they are of low intelligence.

The Littleton House subjects frequently omitted questions in Part I though, with the exception of No. 11, they never omitted questions in Part II. The W.E.A. group always took the questions in order, on both Parts of the test.

The number and type of mistake made by the mentally-defective boys was compared with that made by the W.E.A. students. Apart from the fact that the boys made far more mistakes than the students, the most striking difference lay in the consistency of the errors. The students had, with rare exceptions, stuck to the right answer once they had achieved it, and they tended to give the same wrong answer to questions they failed to understand. The Littleton House boys, however, would often give a wrong solution to a question they had previously answered correctly, and would very often give a series of wrong answers, week after week—changing from one incorrect solution to another. This would suggest that their results were due entirely to chance, but the Analysis of Variance did suggest some measure of genuine improvement.

FIGURE 5 (a)

CORRELATION BETWEEN PARTS I AND II—LITTLETON HOUSE GROUP

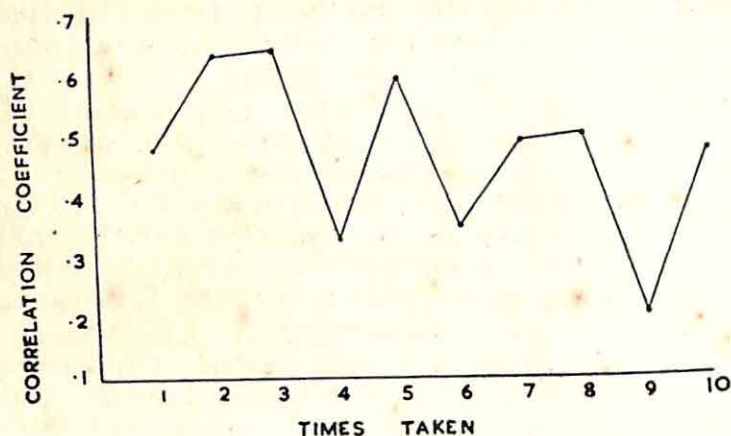
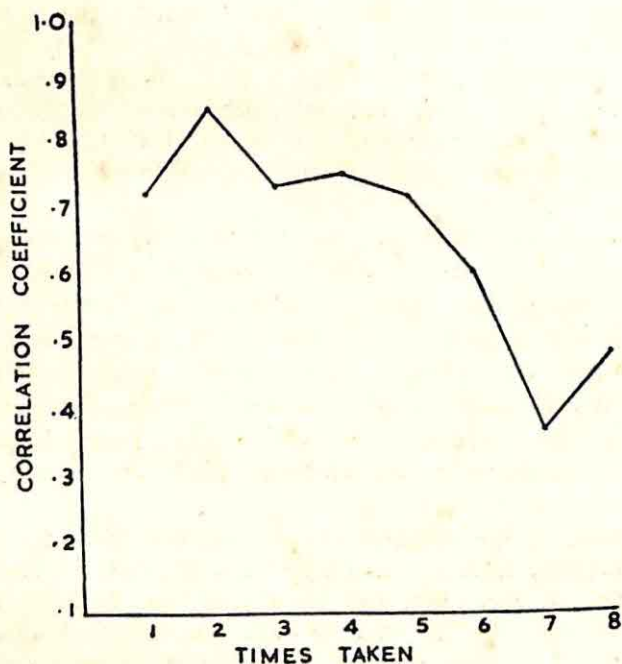


FIGURE 5 (b)

CORRELATION BETWEEN PARTS I AND II—W.E.A. GROUP



IV

DISCUSSION

If the mere taking of one test ten times has the effect of lifting a group of high-grade mental defectives well into the score range of a group of unselected subjects, it seems likely that with knowledge of results and with suitable explanation still more startling effects may be produced. It would be interesting to do further work along these lines, giving positive training, in addition to experience, in test-taking.

The present findings have immediate bearing on two practical problems: the weight to attach to psychological test scores when allocating to schools normal children who may or may not have been previously coached for such tests and the extent to which mental defectives can learn from experience and perhaps compete successfully, in certain fields, with more fortunate people.

It is evident that practice effects play as large a part in intelligence test taking as in many other skills, in spite of the fact that these tests are designed to assess *potential* ability—and do so fairly successfully when given to a group for the first time. The results found in this enquiry strongly suggest that the flattening of the “learning curve” observed with the W.E.A. students was due to the ceiling artificially imposed: it is probable that a test which offered the latter more scope for improvement would produce a rise in score comparable to that gained by the mental defectives.

The consistency of the group as assessed by comparing each testing with every other and with the pool of the scores, confirms the results found on this issue with the W.E.A. group. This has no bearing, of course, on the *validity* of the test, nor does it affect the unfortunate situation in which some members of a test group are more test-sophisticated than others. It is necessary for the satisfactory practical application of psychological tests though it is obviously very far from being sufficient.

The ease of Part II as compared with Part I is, in the writers' experience, largely a function of the low intelligence of the group chosen for the present enquiry. It is true that the mean of Part II is, in general, higher than the mean of Part I, but there is some evidence that this discrepancy tends to increase as the intelligence of the group decreases. Children gain markedly higher scores on the perceptual than on the verbal and numerical Part. It is not clear, however, whether this is related to their possession of greater powers of visual imagery than adults, or their lesser intelligence.

With unselected subjects there is a correlation of the order of 0.8 between the two Parts, some subjects doing better relatively on Part I and others on Part II.

The inability of the mental defectives to cope with Part I of AH 4 when contrasted with their facility on Part II and on the Matrix Test, emphasizes the importance of the medium used when assessing the mental ability of low-grade people—it being assumed that the psychologist aims to discover “how good” the subject is rather than “how bad” he is. At higher levels this finding suggests—as other findings have done—that an intelligence test which confines itself to one bias only may yield highly misleading results.

The upward trend of the progress curves suggests that in any case subjects with low intelligence may take a great many trials to reach the level beyond which they fail to improve. It was a pity that the end of term brought testing to a close with the eleventh session. On the existing data it is difficult to determine whether this slow but continuing progress is a property of low-grade groups on an intelligence test or whether the same result would be found with a brighter group on a more difficult test. The same doubt attached to the question whether more intelligent subjects will always improve to a greater extent than the less intelligent when the test is sufficiently difficult to allow full scope to the people originally gaining the higher scores.

Light may be thrown on some of these questions by repeating the experiment with groups of varying levels of intelligence and with a more difficult intelligence test. It might be of interest, too, to repeat the experiment with psychological tests other than those of intelligence.

V

SUMMARY

(a) Twelve mentally-defective schoolboys between 14 and 16 years of age took test AH 4 once a week, ten times. They took also Progressive Matrices once, after they had finished the AH 4 testing.

(b) In addition to the above test scores, the writers had access to the boys' Binet I.Q.'s gained some years previously, and personal assessments from the Headmaster. Examination of these gave inconclusive results and they have, therefore, been omitted from this paper.

(c) The mean progress curves of the subjects showed (i) a tendency for improvement to continue throughout the ten testings, and (ii) marked differences between Part I (verbal and numerical) and Part II (diagrammatic) of the test, the latter starting at a much higher level and showing greater improvement.

(d) Inter-correlations between the testings were highly significant, surprisingly so in view of the selected nature of the group and the erratic nature of the progress of individual subjects. The consistency dropped strikingly, however, when the eight subjects with the lowest scores were treated as a separate group.

(e) The scores of the group on Progressive Matrices were startlingly higher than they would have been expected to be, had the boys taken this test *before* having taken AH 4 repeatedly, and without having completed examples prior to the test proper.

(f) The main differences between the test results of the group of mentally-defective schoolboys and the group of intelligent adults were as follows: (i) The schoolboy's individual progress curves were erratic and apparently unpredictable, whereas the W.E.A. students produced smooth curves with a well-marked trend. (ii) The low-grade group found Part II very much easier than Part I, judging from the number of questions attempted, the ratio between number attempted and number correct, and the number of questions omitted. The W.E.A. group had tended to maintain a higher ratio of correct to attempted on Part I than on Part II, throughout the testings, and these subjects had not omitted questions in either Part. (iii) The individuals in the duller group were far less consistent, both in their errors and in their correct answers, than the individuals in the superior group.

(g) It was suggested that the experiment be repeated with groups differentially pre-selected with respect to intelligence and with tests of different levels and media.

VI

REFERENCES

1. HEIM, A. W., and WALLACE, J. G. (1949). The effects of repeatedly retesting the same group on the same intelligence test: I. Normal adults. *Quart. J. Exp. Psychol.*, 1, 4.
2. WALLACE, J. G. (1949). Results of a high-grade intelligence test applied to a university population. *Unpublished*.

The writers wish to express their warmest thanks to the Headmaster of Littleton House School for granting them facilities for carrying out the work, and for his kindness and help throughout the investigation. They are grateful, too, to Mr. E. G. Chambers for his assistance on the statistical side, and to Miss V. Horn for her help with the figures. They are especially indebted to Miss V. R. Cane and Dr. and Mrs. A. D. Harris for their invaluable help in the interpretation of results.

(Manuscript received 3rd July, 1949)

THE THRESHOLD FOR SUDDEN CHANGES IN THE VELOCITY OF A SEEN OBJECT

BY

W. E. HICK

(From the Medical Research Council Applied Psychology Unit, Cambridge)

The difference threshold for the velocity of a seen object was measured by the method of constant stimuli, using two categories. An approximate correspondence with Weber's law was found, the divergence from it appearing, in general, as an increase of the threshold at both ends of the range of initial velocities. The Mean Threshold (0.5 probability of perception, corrected for guessing) was, in favourable conditions, about 12 per cent. of the initial velocity. Whether the stimulus was an increase or a decrease of velocity made no marked difference. With two moving objects, which converged, crossed, and then diverged, both suffering the same change of velocity, the threshold was higher.

Velocity changes as low as 2.5 per cent. elicited a significant proportion of correct responses. Some theoretical points in connection with this are discussed. Responses to blank stimuli showed a strong tendency to guess "slower," which tendency differed significantly in degree between most of the experimental conditions. Tests with reduced exposure times showed that exposures could be as short as 0.5 second (the velocity change occurring in the middle of the exposure) without appreciable detriment.

I

INTRODUCTION

THE estimation of rates of change of quantities is an essential part of many forms of prediction. In a number of important cases the quantities are perceived visually, and are of the nature of positions or shapes of objects. The particular task which suggested this study is that of tracking, in which, by the operation of a control handle and its connected mechanism, two objects are required to be kept in coincidence in spite of a continuous tendency to diverge. Other tasks, similar in principle, and on which the present findings may have a special bearing, are driving in convoy at night, and flying in formation. Many more examples could be given, the essential common feature for this purpose being that the primary stimulus is a spatial dimension and that efficient control demands anticipation of its future value on the basis of its present rates of change—in other words, the ability to judge velocity and, if possible, higher derivatives.

Little previous work seems to have been done on this subject. Dimmick and Karl (1930) measured the effect of exposure time on the least perceptible velocity, and other investigators have studied the influence of illumination, position in the visual field, background, etc., but no reference has been found dealing with the ability to discriminate between different successive velocities of a moving object. This is essentially the subject of the present study. It can be regarded as a measurement of the difference threshold for velocity, or as a preliminary approach to the measurement of the acceleration threshold.

II

METHODS AND APPARATUS

Two methods, using quite different apparatus, were employed. In the first (called the "drum" method) an endless band of paper some 6 ft. long passed over a drum, in front of which was a mask having a narrow slit. The moving object was provided by a line on the paper, a short segment of the line being visible through the slit. The motion of the object was horizontal, and consisted of periods of between 2 and 4 seconds of constant velocity, the changes of velocity being instantaneous.

The subject was seated with his eyes approximately 18 in. from the mask. He was provided with a knob, to be moved to right or left according to the direction of the momentary acceleration. The instruction was not to hold the knob over, but merely to apply a brief pressure in the correct direction as soon as a change of velocity was perceived.

In the first trials the object was liable to accelerate or decelerate when moving either to the left or the right, and a deceleration would occasionally be sufficient to cause a reversal of motion. In order to explain the task to the subject he was asked to imagine that the object, when it changed speed, had been given a push, and he was to press the knob in the same direction as the imaginary "push." Although a few subjects grasped the requirement at once, and most did so after some practice and demonstration, it was soon evident that the unfamiliarity of the concept of positive and negative acceleration was causing errors that had nothing to do with perception. The task was therefore simplified by making the spot move only from right to left, with no reversals. As soon as it had reached the left-hand end of the slit it began again at the right-hand end.

Even so, this method exhibited certain defects in practice, the chief of which was that, in order to get a measure of the reaction time, it was necessary that the stimuli should appear at irregular intervals. The result of this and other factors was that stimuli many times greater than threshold value were frequently missed, which led to complications in assessing the results. The technique of "forced guessing" could not be used, and therefore there was no control over the degree of confidence demanded by the subject before he would respond. Moreover, it was impracticable, with the drawing instruments at hand, to produce a standard set of stimuli in sufficient numbers to prevent the sequence on each band being liable to be learnt. The lines were, in fact, drawn haphazard, within general limits, and this prevented learning.

The second method (called the "oscilloscope" method) was devised to reduce these defects. To a standard Furzehill oscilloscope with D.C. amplifiers was coupled what was essentially a Miller circuit. By means of a multiway switch, condensers could be selected to give sweep speeds of the spot of 1.32, 3.12, 7.02, 11.4, 22.98, 49.5 and 89.3 mm. per second on the screen. These were measured by photographing the spot on moving film, and the constancy of the speed across the tube face was checked by the same means.

The circuit was so arranged that extra resistors could be connected in parallel with that through which the condensers charged, and these, by changing the rate of charging, changed the speed of the spot. The velocity increments obtainable in this way were fixed percentages of the initial velocity, and ranged from 2.5 to 75 per cent. for increases and from 2.44 to 42.9 per cent. for decreases. These percentages were checked partly photographically and partly by stop-watch (using the lower initial speeds).

The switching in of the parallel resistance was performed automatically when the spot reached a predetermined position. A silent switch was provided, out of sight of the subject, for the experimenter to pre-select "faster" or "slower." This automatic switching device, which consisted essentially of a triode and two relays, was given a separate H.T. supply, to eliminate any possible interference which might have jerked the spot at the moment of operation.

In order to extend the range of the experiments two further elaborations were made. One was to produce the appearance of two spots by means of a commutator, driven by a motor at well above flicker speed, which repeatedly reversed the input to the oscilloscope from the Miller circuit. The two spots approached one another from opposite sides of the tube and crossed in the middle. The other was a pair of shutters which could be placed so as to cover all but the middle portion of the tube face, the aperture being adjustable as required. The purpose of this was to examine the effect of altering the exposure of the spot. Two difficulties were encountered here. The spot-position at which the automatic switch operated was subject to random variations from occasion to occasion, with a standard deviation of about 0.9 mm., not to mention a slow drift, and therefore it was not practicable to reduce the aperture to less than 1 cm. In addition, it was clearly essential to give the subject proper warning of when the spot would appear; this problem was solved by incorporating in one of the shutters a piece of transparent plastic, moulded so as to break up the light; through this a faint irregular glow from the approaching spot could be seen, but no clue as to its velocity could be gained.

The experimental procedure with this apparatus (the oscilloscope) was as follows. The subject was seated with his eyes at approximately 21 in. from the display, and told to say "faster" or "slower" according to his judgment of the change of speed. Reaction

time was not, of course, recorded. At the beginning of a series he was told approximately where on the tube the change of speed would occur, and that he would know by the click of the relays each time that it *had* occurred; conditions were therefore as favourable as possible for expectancy and attention.

The subject was also told, untruthfully, that there would always be *some* change. These blanks amounted to about 15 per cent. of all stimuli (except in the case of the reduced-aperture experiment, where they were not used at all), and, as will be seen later, they contributed an indispensable item of information.

The general procedure was to begin by giving the subject a few practice trials, and then to present the standard series of stimuli in descending, and then ascending, order of magnitude, at each value of initial velocity. In fact, this rule was not strictly adhered to, because it was thought that, since the method of forced guessing (no third category) was being used, a considerable number of not consciously perceptible stimuli in succession might have a bad effect on motivation; for this reason a relatively easy stimulus was always introduced after three or four "impossible" ones. The series of initial velocities was only given once to each subject, at least with the same conditions, but approximately equal numbers received it in ascending and descending orders.

Six subjects were used in the drum method and eighteen others with the oscilloscope, but not all the eighteen were tested in all the conditions. The conditions not already mentioned refer to the positions of the velocity change in the case of the double spot. These were:—

- (1) Spots 2 cm. apart, converging.
- (2) Spots 2 cm. apart, diverging.
- (3) Spots just crossing each other.

III

RESULTS

The definition of the threshold which was chosen was that of the so-called Mean Threshold. It is that value of the stimulus which makes the number of smaller stimuli perceived equal to the number of greater stimuli not perceived. It can be given a physical meaning by assuming that there is, at any instant, a definite and absolute threshold, but that it varies from instant to instant; then the Mean Threshold is the *median* value of the instantaneous threshold. Consequently it is an estimate of the stimulus value which has a 0.5 probability of being perceived, notwithstanding that this stimulus may, in the experiment, have been perceived on much more or less than 50 per cent. of occasions. Such a discrepancy has then to be ascribed to sampling error.

This method has the advantage of making some use of all the data, and of reducing the labour of computation to the realm of practicality; indeed, where the stimuli are separated by equal intervals and are presented equal numbers of times, the work amounts to little more than counting the total number perceived. Unfortunately these conditions could not be satisfied in the present case.

On the other hand, it is essential to the method that the stimuli should cover the whole transition region; that is, they must range from the "always seen" to the "never seen." The data in the present study required a certain amount of manipulation to bring them into reasonable conformity with this requirement. In the case of the drum method, even the largest velocity changes used (up to about thirty times the threshold value) were occasionally missed. In fact, above the apparent transition region the probability of seeing a stimulus could be seen to have little, if any, relation to its magnitude. Doubtless the reason was something like inattention or confusion. These probabilities were estimated (they were of the order of 0.9), and it was assumed that the probability of actually responding to a given stimulus was the product of the probability of its being above the instantaneous threshold and

that of the subject being in a suitable state to respond to any stimulus. Thence a correction to the observed probabilities over the whole range of stimuli (for each initial velocity) was obtained, having the effect of increasing them by some 10 per cent. (except, of course, those that were already unity).

In the oscilloscope experiments it was necessary to correct the observed frequencies for chance guessing. The responses to the blank stimuli were used for this purpose; they showed a strong over-all preference for "slower" rather than "faster," to the extent of nearly 75 per cent. in one particular condition. Had the conventional assumption of an equal chance of guessing one or the other been used the results would have been grossly distorted.

With the two-category technique it is, of course, impossible to give a stimulus so small that it can be said with certainty to be never seen. The correction for chance guessing can, at best, only be right on the average. Taking all the oscilloscope data for the smallest velocity change (2.5 per cent.), an appreciable but not significant preponderance of correct reports was found. There were several individual cases of highly significant frequencies of correct responses to this small stimulus, but that is natural in any large collection of data, and need not indicate any real effect. However, the unavoidable result was that the corrected frequency of perception of the smallest stimulus was sometimes positive and sometimes negative, but seldom zero; hence the conditions for computing the Mean Threshold by the method chosen were not exactly satisfied. The thresholds, as finally calculated, are therefore probably a little too large, but the error is certainly very small.

Theoretically, it is possible to estimate the standard error of the Mean Threshold, since the variable term in the formula is essentially the sum of a number of binomially-distributed quantities. But it appeared that there were difficulties in applying the method usefully to the present case, and it was thought that, as a number of thresholds were calculated, the variability would be sufficiently evident for practical purposes.

The Mean Thresholds obtained with the "drum" method are given below (Table I), in percentages of the initial velocities. The latter are in degrees per second at the eye, and are group means.

TABLE I

Initial velocity	0.11	0.38	0.77	1.28	2.31	4.18
Mean Threshold (per cent.)	133	41	46	31	21	12

The only interest lying in these figures concerns their marked disagreement with those obtained by the oscilloscope method, which are given below (Table II). It can be seen that the tendency to miss perceptible stimuli is very great at the lowest speed, but declines rapidly as the speed is increased, until, in the region of 4 degrees per second, it is inappreciable. The reason is probably that, at the higher speeds, the spot made a succession of fairly rapid sweeps across the display, and the subject properly inferred that a velocity change would occur somewhere in each sweep. Consequently he had a clear warning signal, and, in addition, the circumstances would favour an attitude approaching that of forced guessing. At the low speeds, on the other hand, the spot drifted slowly across, suffering a number of velocity changes in its course, and the subject might be both less attentive at the relevant instants and more inclined to wait for an unmistakable stimulus before responding. This was one reason for repeating the experiment with the oscilloscope apparatus.

With the drum method there were 29 responses in the wrong direction out of a total of 962 (3 per cent.). A significant majority of these wrong movements were in response to the second of a pair of stimuli, both of the same direction. Since there was only one case of two wrong responses in the same direction, it is likely that the principal cause of wrong responses here was a tendency not to do the same thing twice in succession, as found by Thorndike (1927) and Telford (1931) among others.

TABLE II

MEAN THRESHOLDS AS PERCENTAGES OF INITIAL VELOCITIES, FROM OSCILLOSCOPE DATA
(Degrees per second at Eye)

<i>(a) Single Spot: Velocity Increment.</i>								
Initial velocity	..	0.15	0.35	0.81	1.31	2.64	5.69	10.25
Mean Threshold	..	12.6	9.9	9.0	9.6	9.6	16.0	13.2
<i>(b) Single Spot: Velocity Decrement.</i>								
Mean Threshold	..	15.0	14.7	14.4	13.5	9.3	12.2	20.0
<i>(c) Double Spot (Converging): Increment.</i>								
Mean Threshold	..	17.5	15.2	11.4	8.3	7.2	11.7	20.8
<i>(d) Double Spot (Converging): Decrement.</i>								
Mean Threshold	..	14.6	13.8	17.9	18.1	19.9	19.3	19.1
<i>(e) Double Spot (Diverging): Increment.</i>								
Mean Threshold	..	22.1	—	21.2	—	14.9	—	19.3
<i>(f) Double Spot (Diverging): Decrement.</i>								
Mean Threshold	..	12.7	—	13.4	—	11.4	—	12.9
<i>(g) Double Spot (Crossing): Increment.</i>								
Mean Threshold	..	17.1	—	13.3	—	13.7	—	10.7
<i>(h) Double Spot (Crossing): Decrement.</i>								
Mean Threshold	..	15.9	—	13.6	—	10.6	—	10.9

Table II shows that the Mean Threshold obeys Weber's law approximately, the principal divergence from the law appearing, in most of the conditions tried, as an increase of the threshold at both ends of the range of initial speeds. This very common finding is clearly inevitable in the present case if the range is extensive enough. Very slow motion, as of the hand of a clock, ceases to be appreciated as such, and the only perceptible change of velocity is one which raises the velocity to the level of being detectable as velocity, and this could be a very large relative increase. Decreases are only perceptible in terms of remembered displacements and time intervals, one or both being necessarily liable to large error. Very fast motion likewise ceases to be appreciable as motion. It is true that, in these experiments, the extremes were not reached, but it appears that they were approached. At the lowest speed the spot did appear motionless at a casual glance; while, at the highest speed, the limitation may, as discussed below, have been the short time during which the spot was in view.

The results from the single-spot experiment indicate that the threshold is in the region of 12 per cent. of the initial velocity; this is confirmed by the appropriate portion of the data from the subsequent reduced-aperture tests, and may, therefore, be taken as reasonably reliable. With the double spot, however, the results in the

various conditions are somewhat divergent, and it is difficult to know how to interpret them. There is some general tendency for the thresholds to be higher than with the single spot; in this connection it is interesting to note that most subjects concentrated on one spot, the other being regarded as more or less confusing and distracting. To have looked at both would have been, in effect, to employ parafoveal vision, which is usually said to be more sensitive to movement than foveal vision; but the latter may be found to be a better detector of change of velocity—in other words, a better measurer of a velocity that is well above threshold in absolute magnitude. The present results only raise this question, of course; they do not answer it.

A clearer picture of the effects of the different conditions is obtained when the figures for increments and decrements are pooled, as in Table III.

TABLE III

	Initial velocity	..	0.15	0.35	0.81	1.31	2.64	5.69	10.25
(a) <i>Single Spot.</i>									
Mean Threshold	..	13.8	12.3	11.7	11.6	9.5	14.1	16.6	
(b) <i>Double Spot—converging.</i>									
Mean Threshold	..	16.1	14.5	14.7	13.2	13.6	15.5	20.0	
(c) <i>Double Spot—diverging.</i>									
Mean Threshold	..	17.4	—	17.3	—	13.2	—	16.1	
(d) <i>Double Spot—crossing point.</i>									
Mean Threshold	..	16.5	—	13.5	—	12.2	—	10.8	

The probabilities of guessing "faster," as estimated from the responses to the blank stimuli, were, for the various conditions:—

(a) and (b)	(c) and (d)	(e) and (f)	(g) and (h)
0.32	0.45	0.25	0.46

Of these, 0.45 and 0.25 differ significantly at the 0.5 per cent. level, and the other possible pairs (except 0.45 and 0.46) also differ significantly, though less so. Since the same subjects were used in conditions (a) to (d), and some of them performed in the other conditions as well, the differences cannot be attributed to personal idiosyncrasies. The cause must presumably have to do with the different conditions, but what it is can only be conjectured. Naturally, when a subject is asked to arrive at a decision, all rational grounds for it having been carefully removed, he must base it on irrational grounds—on such symbolic suggestions as he may be able to glean from the total situation. It is always of interest when such suggestions are found to operate in the same way on the majority of subjects. It may be noted that 289 of the total of 805 blanks were guessed as "faster," which is so much less than the pure chance expectation as to be far beyond the range of the usual table of χ^2 . Yet the 2.5 per cent. increase of velocity was correctly reported 98 times out of 184, which indicates, with a significance better than 0.1 per cent., that even this small change is capable of reaching the cerebral apparatus.

IV

DISCUSSION

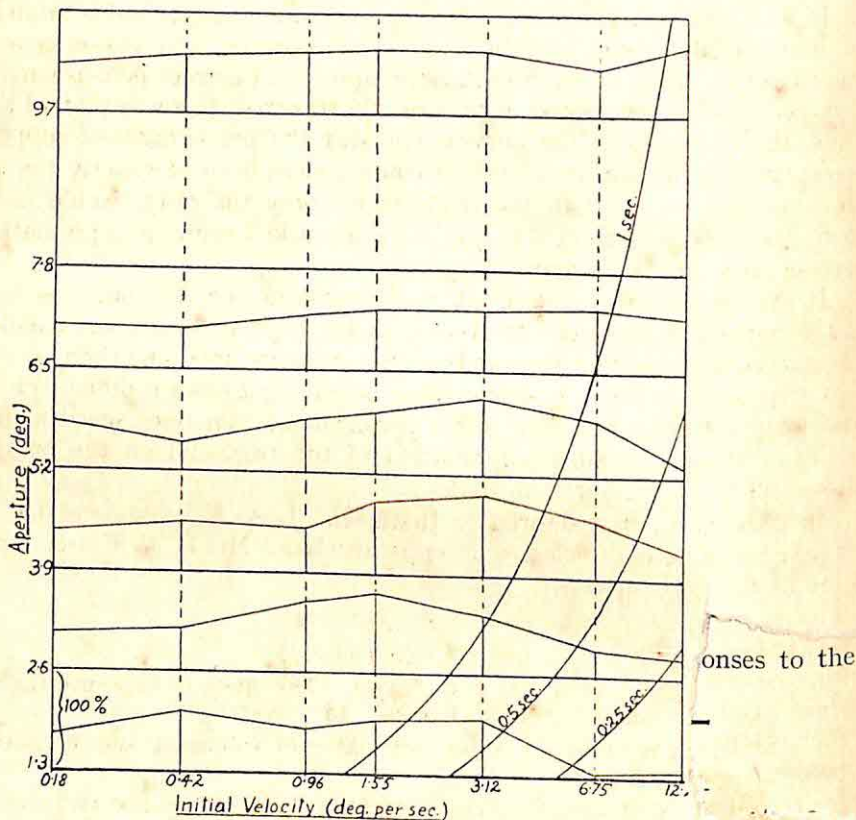
These results illustrate the advantages and disadvantages of the two-category or forced-guessing method. It eliminates the very uncertain factor of the degree of confidence the subject requires in his judgments before he will admit that they are "conscious." It would be better if we could control, or at least measure, this factor. But eliminating it has the justification that many of the judgments made in the course of a skilled activity are not conscious; in fact, the circumstances must often amount to a requirement for forced-guessing. The over-all performance must be considerably affected by judgments which, if they were made in isolation in experimental conditions, the individual would have no subjective confidence in, but which, on the whole, are more often right than wrong.

Then, behind this, there is the prevalent notion that there is a real physiological threshold, towards which observed thresholds tend with long practice. Many elementary physiological experiments and demonstrations have contributed to this notion, which implies that the nervous system is a true ankylotic system, with something approaching an absolute discontinuity between adequate and inadequate stimuli. On the other hand, experiments of the kind described here suggest that, if there is a definite physiological threshold, it is sometimes very far below what would normally be called the psychological threshold. It is possible that even the smallest stimulus, provided it has a real physical existence—i.e. contains at least one quantum of energy—is capable of producing a bias towards making the correct response. The bias may be extremely small, and yet produce a significant effect in a long enough series of trials. It has been argued by de Vries (1948) that sensory cells and nerve fibres are as sensitive as it is possible for them to be without suffering an undue number of spontaneous excitations. If this is so—or, indeed, if it be accepted that spontaneous excitations, due to random processes at the molecular level, do occur—it follows that a stimulus too small to discharge a cell when it is in its average condition will, nevertheless, slightly increase the probability of excitation. It is evident, of course, that even if a sensory cell succeeds in discharging, the resulting signal has to run the gauntlet of many other randomizing processes before it eventually finds expression in activity.

It remains an open question why a large or relatively unmistakable stimulus gives rise to a conscious sensation of having responded correctly, while a very small stimulus, even when it evokes the correct response, leaves the subject feeling that he has made a pure guess. A reasonable hypothesis seems to be that the deciding factor is immediate memory; if the subject's memory image of the stimulus suggests to him the same response as the one he actually made, he may fairly conclude that he was right; but a very small stimulus may produce a trace which is obliterated almost at once, leaving the subject unable to decide whether he really perceived it or not. The alternative possibility that the feeling of confidence does not have to await this review of the memory image, but is born with the response, and that the image is only a kind of rationalization projected into the past, is probably only an apparent alternative—another way of saying the same thing. The fact that memory images in general can be searched for details not originally asked for does not help, because it can always be argued that these details caused changes of attitude or "set" which, though not visible, were just as much entitled to be called responses as the overt movement. Although the writer believes that this kind of argument can be fitted into a consistent system, the cornerstone of which is the thesis that knowledge is fundamentally operational in character, it is unnecessary to pursue

the matter here; in any case, the question raised is possibly not one that experimental psychology, as such, can answer. Even the fact that, in the present experiments, "easy" stimuli were usually responded to in a loud and cheerful voice, contrasting with the dull monotone evoked by the more difficult ones, means little more than that different stimuli produced different responses.

FIGURE 1



With regard to the reaction time records, the principal finding was the usual one—that the more difficult the stimulus the longer was the reaction time. As explained above, the reaction times were calculated from the "drum" data, only those for the correct responses being taken account of. On the whole, the times were longer for the initial velocities, which accords with the higher thresholds observed in these conditions. Mean reaction times varied from 2.1 seconds to 0.38 seconds. The actual figures do not appear to convey any important information, and are therefore omitted.

The essence of the results obtained in the reduced-aperture experiment is given in the diagram. The velocity change was set at approximately 12 per cent., this being about the value which would result in something like 50 per cent. of correct perceptions—i.e. it was near the Mean Threshold—and would therefore be expected to yield the most information. Eight subjects were used, and the technique was that of forced-guessing, as before. Starting with the largest aperture the subjects were given equal number of increments and decrements at each initial velocity, starting with the lowest; in other words, the initial velocities were presented in

ascending order, and repeated at each aperture, the latter being in descending order. Theoretically, a fully randomized order would have been better, but there were several practical objections. In the first place all these tests were monotonous and fatiguing to the eyes, and the forced-guessing procedure is an unsatisfying one to the subject; from preliminary trials it seemed that a regular progress from the easier to the more difficult was the least disturbing to the subjects, and the most likely to secure their co-operation.

It was assumed that, since incremental and decremental stimuli were pooled, random guessing would be unbiased. The excess of right answers over 50 per cent. was therefore taken as representing the number of correct perceptions; for example, if 75 per cent. of the stimuli were correctly reported, this was plotted as 50 per cent. Thus, the ordinates of the curves represent the percentages of supposedly correct perceptions. The curves would, of course, have been of exactly the same shape if the data had been left in the crude state; only the range would have been from 50 to 100 instead of from 0 to 100. Each marked ordinate represents a percentage derived from about 60 actual responses.

It can be seen that the effect of the reduced aperture only becomes noticeable at the bottom right-hand corner—i.e. at the highest speeds and smallest apertures. The curved lines marked in seconds are iso-exposure lines, and there is some suggestion that exposures below 0.5 second were beginning to cause difficulty. It is evident that aperture alone does not affect performance; nor does speed alone. Exposure is the most likely simple combination of the two, and on the present data it is impossible to go further than that.

In conclusion, it is desired to thank Dr. J. A. V. Bates and Dr. A. Carpenter for assistance with the electronic apparatus, and Mr. D. C. Fraser for carrying out some of the earlier tests.

V

REFERENCES

- level, do occur in its average motion. It is the resulting motion. *J. Exp. Psychol.*, **13**, 365-369.
- before it ever remains, E. L. (1927). The refractory phase of voluntary and associative responses. *Psychol.*, **14**, 1-36.
- It remains, E. L. (1927). The refractory period of associative processes. *Psychol.*, **14**, 234-236.
4. DE VRIES, H. (1948). Minimum perceptible energy and Brownian motion in sensory processes. *Nature*, **161**, 4080, 63.

(Manuscript received 28th September, 1948.)

PROCEEDINGS OF THE EXPERIMENTAL PSYCHOLOGY GROUP 1948-9

- April 24-25, 1948.* Extended Meeting at Cambridge. *1st Session:* Symposium on Artistic Expression in Relation to Mental Disorders, by E. B. Davies (by invitation), G. Westby (by invitation) and R. W. Pickford. *2nd Session:* "Parallels between the Conditioned Reflex and the Phi-Phenomenon," by George Humphrey (by invitation). *3rd Session:* "Experiments on Repeated Testing of Intelligence," by A. W. Heim. "Preliminary Studies of the Changes in Skilled Performance associated with Increasing Age," by A. T. Welford.
- July 24, 1948.* Social Meeting in Edinburgh for Members and Guests attending the 12th International Congress of Psychology.
- December 4, 1948.* (London): "Some Experiments on the Visual Perception of Size," by R. B. Joynson. "Non-Verbal Reasoning in Cases of Cerebral Lesion," by A. E. Sladden.
- January 8, 1949.* 2nd Annual General Meeting in London. "The Functional Significance of Electrical Systems in the Brain," by W. Grey Walter (by invitation).
- March 30-31, 1949.* Extended Meeting at Oxford. *1st Session:* "Form Perception in Dark Adapted Vision," by B. Semeonoff. *2nd Session:* Symposium on Prefrontal Leucotomy and Related Problems, by C. W. M. Whitty, P. Glees and MacDonald Tow (all by invitation). *3rd Session:* "Experimental Sensory Dissociation," by D. C. Sinclair (by invitation). "Apparent Fluctuations in a Sensory Threshold," by R. C. Oldfield.
- July 23, 1949.* Meeting at Cambridge in association with Symposium on Animal Behaviour arranged by the Society for Experimental Biology and the Institute for the Study of Animal Behaviour. *1st Session:* "Researches at the Yerkes Laboratory of Primate Biology," by K. S. Lashley (by invitation). "Perception and Insight," by K. S. Lorenz (by invitation). *2nd Session:* Short interim reports of work in progress at the Cambridge Psychological Laboratory, by Nancy Harris, P. H. R. James, C. Poulton and C. B. Gibbs (all by invitation).
- September 15-16, 1949.* (Oxford): *1st Session:* "Work in Progress at the Psychological Department of the National Hospital, Queen Square: An Informal Review," by J. McFie (by invitation), M. F. Piercy (by invitation) and O. L. Zangwill. *2nd Session:* "Cybernetics: How Nervous Structures have Ideas," by W. S. McCulloch (by invitation), followed by Discussion. *3rd Session:* "Degrees of Assurance," by B. Babington Smith.

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Vol. II

1950

Part 2

THE EFFECT OF ONE EXPERIENCE UPON THE RECALL OF ANOTHER

BY

D. RUSSELL DAVIS and DURGANAND SINHA

(From the Psychological Laboratory, University of Cambridge)

An exploratory experiment was made to show how one experience may exert an influence upon the recall of another, and how both may lose their separate identities in memory, to become merged into what Bartlett has described as an active organization of related experiences. The original material was a story, and the experience interpolated between it and its recall was a picture which illustrated part of it, in some respects correctly, in others incorrectly. Some subjects were asked to recall the story, others the picture, after from one to four weeks, or in a few cases a year.

The picture was viewed under the influence of the attitudes induced by the story. These determined its selection from a number of other unrelated pictures, to be perceived as a related experience. In consequence only few details gained attention, and these were interpreted in terms of the attitudes induced by the story, in some cases contrary to the meaning given to them by the rest of the picture. Other details without significance in the theme of the story were neglected. In its turn, seeing the picture favoured the recall of some aspects of the theme of the story, details not reinforced by the picture tending to be left out in recall. Many picture details *intruded* into the reproductions of the story. This tendency was progressive, and at the end of a year picture and story details were not separated. Especially, the picture brought about changes in the points of emphasis in the story.

I

INTRODUCTION

THE essence of Bartlett's theory of remembering is the idea that past experience operates as "an organized mass rather than as a group of elements each of which retains its specific character" (Bartlett, 1932). Experiences are retained in memory in the form of active organizations or schemata, a schema being a reification of experience conceived of as a plastic model which is reshaped by every new experience related to it. Schemata may be partially independent of one another, but tend to overlap and fuse.

The several experiments which he describes justify Bartlett's view of memories as constructions of related experiences undergoing continuous modification. Yet it cannot be denied that the memories of some experiences remain discrete, or at least relatively independent of one another, fixed and uncontaminated, as, e.g., the theory of Freud requires; for even if the clinical evidence is discounted—and this would hardly be reasonable—there are in everyone's personal experience impressive examples of events, usually of some special significance, which have resisted assimilation into general experience, and some aspects of which remain capable of relatively accurate or "literal" recall even after the passage of many years.

The theories of Bartlett and Freud are not incompatible with one another in this respect. Indeed, although this point has often been missed, Freud (e.g., 1892, pp. 29-34) emphasized the special character of the memories which give rise to hysterical phenomena. Thus he remarked upon their "wonderful freshness," "astonishing integrity and intensity of feeling" and "their exceptional fate in reference to all the ordinary processes of effacement," and suggested that they are "preserved with such freshness and affective force because the normal process of absorption by abreaction and by reproduction in a state of unrestrained association is denied them." On the other hand, he argued, the more usual process is for a memory to become "merged into the great complex of associations," and then to become "ranged alongside of other experiences which perhaps contradict it; thus it undergoes correction by means of other ideas."

It is pertinent therefore to enquire further into the conditions which favour the preservation in memory of the separate identity of an experience, or its merging into an "active organization" or "complex of associations" of related experiences. Before the attempt is made to overcome the difficulties of a direct experimental approach, however, the problem requires clearer definition, and in our opinion there is a need for a number of exploratory experiments. Accordingly, we have made several experiments to test the influence of interpolated material of various kinds upon the reproduction of an original experience. The present paper describes a study of this influence when the original material was a story and the interpolated material a picture which illustrated part of the story (though not being in complete agreement in every detail). Some subjects were asked to recall the story, and others the picture. The investigation was intended to throw light on the manner of recall and the kind of items recalled.

II

MATERIAL AND PROCEDURE

The story and picture were so chosen that the picture was relevant to, and could be expected to exert a big influence upon the recall of the story. The picture was a coloured postcard reproduction of Pieter Breughel's painting, "The Village Wedding," with which, although it is famous, our subjects were not familiar. Around the scene there depicted was written in a pompous style a story of about 750 words, partially descriptive and partially episodic. This story (which is printed as an appendix) gave to the scene of conviviality a significance quite different from that which the painter intended; its first part described a feud between two families and their reconciliation in the betrothal of the son of one with the daughter of the other; its latter part, a wedding feast, some details from the picture being rendered correctly, some being hinted at, some being deliberately misrepresented and others being left out; in this part of the story mention was made of several untoward incidents which nearly led to a revival of the feud.

The subjects were university students. They were divided into four groups, who were treated according to the plan shown in Table I. Of subordinate interest was the

TABLE I
PLAN OF THE EXPERIMENT

		1st day, original experience	4th day interpolated experience	4th to 28th day recall of
Group A (21)	Reading of story	Presentation of picture	Story
Group B (28)	Reading of story	—	Story
Group C (7)	Reading of story	Presentation of picture	Picture
Group D (7)	—	Presentation of picture	Picture

The number of subjects is given in brackets.

comparison of the recall of the picture in Groups C and D, which showed how the story had influenced the perception and recall of the picture. Comparison of the recall by Groups A and B showed how the picture thus perceived then influenced the recall of the story.

In some respect this plan was similar to that of the experiments on the so-called retroactive inhibition of one habit by another. Our main concern was with the influence of the interpolated material upon the recall of the original material, and this may be thought of as akin to the influence of the performance of one habit upon the subsequent performance of another, when the two habits are similar in some respects. Then facilitation (positive transfer) or interference (negative transfer) is looked for. Only one overall measure of the performance has been obtained in the majority of experiments on this topic. In our experiment, however, the recall of picture or story was studied in a more qualitative than quantitative manner, to show in what respects there was facilitation and in what respects interference, and, furthermore, to what degree there was intrusion of the components of the one into the other.

The procedure was rather more complicated than the foregoing description indicates, for "The Village Wedding" was presented with six other unrelated pictures as part of a quasi-recognition test. The reason for this was twofold: to disguise the real purpose of the experiment; to obtain collateral information about the interaction of picture and story. Three or four days after the story was read before Groups A, B and C, each subject of Groups A and C was shown the seven postcards and asked to identify the one which depicted a scene described in the story. All the postcards were reproductions of paintings in colour. Although several of the other postcards showed rural scenes of festivities, every subject picked out "The Village Wedding," as had been expected. He was then given brief opportunity to examine it. The subjects of Groups A and B were then asked to recall the story on one occasion only, but at intervals varying from immediately to more than three weeks. Those of Group C were asked to recall the picture at similar intervals. Those of Group D, before whom the story was not read, were shown "The Village Wedding" by itself and were then treated in the same way as were those of Group C. Throughout, the instructions to the subjects were framed in such a way that, although the immediate demands were clear, there were no grounds for anticipating what would be required of them on later occasions.

III

RESULTS

(i) *Perception of the Picture.*

Before coming to the effects of the picture on the recall of the story, it is necessary to comment upon the effects of the story on the perception and recall of the picture.

The story seemed to create certain expectations which served as a frame of reference for the perception of the picture. Thus when presented with the seven postcards, most subjects looked for a scene of a certain kind only; one subject said, for instance, "I was looking for a feast in a rustic setting." They had a general impression of what the picture was going to be like, and on that basis "The Village Wedding" was often picked out at once by some process of immediate matching; it was recognized at once as being relevant to the story and as complying with expectations. Similarly, several of the postcards were rejected immediately as "irrelevant" or as not having the right "atmosphere." A reproduction of a painting by Toulouse-Lautrec, for instance, was usually rejected quickly, even scornfully, as "too modern" or as "obviously French." Of interest was this reason for rejecting the same picture, "the dress and atmosphere do not fit," for nothing in the story defined the dress directly, although all would agree that the dress in Toulouse-Lautrec's picture would be out of keeping with such a peasant affair as the story described. Sometimes one other postcard was picked out for more careful consideration because it showed the right kind of scene; this was Pieter Breughel's "The Peasant Dance." It was then discarded, either because some detail was

incompatible with the story, e.g. the scene is an out-of-doors one, or because "The Village Wedding" was preferred, this picture containing details which had been mentioned in the story.

"The Village Wedding" having been chosen in this way, on the basis of general impression, subjects then usually confirmed their decision by looking for details mentioned in the story and expressed their confidence as soon as they had found them. Three details especially served this purpose: "pitchers of wine being filled," "crossed sheaves of corn" and "two servants bearing ash staves." These details were mentioned in the story, the latter two having a special significance in its theme. The former two were fairly prominent in the picture. With the last there commonly occurred a false perception; the two servants bearing ash staves being identified in the picture with what the subjects who had not heard the story described more correctly as "musicians" or sometimes as "bag-pipers."

In the majority of subjects, general impression played the primary, and details a secondary rôle in the selection of "The Village Wedding." Details determined the immediate choice in rather less than one-third of our subjects. Even these subjects, however, who employed a distinctly analytic method discarded other pictures on a basis of general impression. Some subjects expected to see certain details, which, as it happened, were absent from "The Village Wedding" and had imagined a scene with different detail; nevertheless they chose "The Village Wedding" and felt little doubt of the correctness of the choice, although they might express annoyance at the discrepancies between expectation and actuality.

Recall of the Picture.—The subjects who had heard the story tended to recall those details in the picture which had played an important rôle in the story, and which were easily located in the picture. On the other hand, details equally prominent in the picture, but not mentioned in the story, were not recalled at all or, if they were recalled, were treated as accessory. Many of these subjects clearly indicated in their reports that once "The Village Wedding" had been chosen, and the choice verified by one or more details, they had hardly bothered to examine the picture carefully. They interpreted the few details of the picture, which gained their attention, ascribing to them qualities derived from the story, e.g. a bridegroom was often identified, "sitting in the place of honour" and "appearing sad and downcast"; and they recalled those details for which the story had provided an interpretation.

Those who had not heard the story tended merely to enumerate the details of the picture, and their reports revealed careful and minute observation of detail. The number of subjects in Groups C and D was too small for quantitative comparisons to yield significant results, but there was little doubt that, considered from the point of view of the number of items recalled, those who had heard the story gave poorer accounts of the picture than those who had not heard the story; i.e., in respect of the number of items, hearing the story was a handicap in the recall of the picture.

(ii) *Recall of the Story.*

(a) *Amount of Recall.*—In order to obtain a quantitative estimate of the amount of recall, the original story was divided up arbitrarily into about 120 items (as has been done in the appendix), and a similar enumeration of items was carried out for each reproduction. Each item in a reproduction was examined with reference to the original story and then classified. The method was crude, but was regarded as sufficient for its purpose. It was found that the reproductions of those to whom the picture had been shown (Group A) contained more items than did those of the

other subjects (Group B). The size of the difference between Groups A and B is shown by the facts presented in Table II. The story was divided into two parts: the antecedent part which comprised the incidents leading up to the wedding feast;

TABLE II
MEDIAN NUMBER OF ITEMS RECALLED IN EACH SUB-GROUP—WHOLE STORY

	<i>Recall after less than 8 days</i>	<i>Recall after 21-28 days</i>
Group A	50	43
Group B	45	24
For the τ correlation, P is ..	0.02	0.005

The method of Whitfield (1947) was used for the calculation of probability, and median scores are cited here, for this method allowed the combination of several sub-groups differing in the interval before recall.

and the main part, the description of the feast itself. As might be expected, the interpolated experience favoured the recall of the main rather than the antecedent part (Table III).

TABLE III
MEDIAN NUMBER OF ITEMS RECALLED IN EACH SUB-GROUP—
(1) ANTECEDENT AND (2) MAIN PARTS SEPARATELY

	<i>Recall after less than 8 days</i>		<i>Recall after 21-28 days</i>	
	(1)	(2)	(1)	(2)
Group A	15	35	14	29
Group B	16	29	8	16
For the τ correlation, P is ..	—	0.01	—	0.001

(b) *Importations*.—The items in each reproduction were divided into two classes: accurately recalled, simply transformed or distorted details, with counterparts in the original; importations, without counterparts in the original. A large part, but not the whole of the difference between Groups A and B in the number of items recalled was found to be due to the larger number of importations in Group A.

TABLE IV
MEDIAN NUMBER OF IMPORTATIONS IN EACH SUB-GROUP

	<i>Recall after less than 8 days</i>	<i>Recall after 21-28 days</i>	<i>For the τ correlation, interval \times number of importations, P is</i>
Group A	4	7	0.01
Group B	3	4	0.05
For the τ correlation, group \times number of importations, P is ..	—	0.01	—

Of these importations the majority were *intrusions*, i.e. details derived from the picture. Intrusions are one of three types of importation, the other two types being inventions and rationalizations. The three types have been treated together in Table IV, from which it will be seen that:

- (i) In recall after less than 8 days the difference between Groups A and B is not significant.
- (ii) In both groups there were more importations in recall after 21 days than after less than 8 days.
- (iii) In recall after 21 days there were more importations in Group A than B.

These results are in agreement with the finding of Bartlett that, "the tendency to invent or to import new material from a different setting may increase considerably with lapse of time."

(c) *Mode of Recall*.—The recall of those who had not seen the picture tended to be of the *sequential* type; i.e. they reported merely a chain of events with a minimum of circumstantial detail. Also, they tended to use condensed or compressed forms of expression; for instance, in describing the main theme of the story, they used such expressions as "regaled with truly magnificent feast," "there was much feasting and merrymaking" and "the scene was a merry one." In contrast, the recall of those who had seen the picture was fuller and more graphic, e.g. "there were two long tables (*intrusion*); at one sat the two families, at the other the visitors (*invention*): "Plates upon plates of food were brought (*accurate recall*): "The servants stood oafishly eyeing the good things before them (*intrusion and distortion*): "with stupid faces and mouths oafishly hanging down (*distortion*)."

In commenting upon their recall of the story, subjects reported what they had considered to be the crucial details, and from what, in the manner described by Bartlett, they had reconstructed the story. Those who had not seen the picture tended to mention as dominant details items from the antecedent part, e.g. the family feud, the straying of the sheep and the death of the father. On the other hand, dominant details were drawn more often from the main part of the story by those who had seen the picture, e.g. the crossed sheaves of corn and the entry of servants bearing staves. Thus the interpolated experience caused a redistribution and reshuffling of the points of emphasis in recall.

(d) *Fate of Certain Details*.—Two details especially were important to the theme of the story and were also prominent in the picture; these were: "crossed sheaves of corn" and "two servants bearing ash staves." The subjects who had seen the picture recalled the former almost invariably, and the latter frequently, but these details tended to drop out of the recall of the other subjects, being recalled after 21 days by less than half and less than one-fifth of them respectively. If they were recalled by these subjects, they were often distorted; e.g. "crossed sheaves of corn" were recalled as "a family crest" and on one occasion even as "on the wall there hung a sheep which everyone suspected was the stolen sheep." Thus some salient details were accentuated by the interpolated experience.

Other circumstantial details, such as "pitchers being filled with wine," which were often remarked upon in the picture, and which contributed in the story only to the atmosphere of feasting and gaiety, were included only occasionally in reproductions of the story. Even then they were remembered as picture details.

Recall After One Year.—A year after the reading of the story the nine subjects of Group A and the seven of Group B who were still available were asked to recall

as much as they could of the story. They were instructed to report what they recalled as it came along, and not to think it over and then present it in a coherent form.

In the number of items recalled there was no longer a difference which could be regarded as significant, but the preferential recall of primary details was still to be noted in those who had seen the picture; for instance, "crossed sheaves of corn" were recalled by all but one of these subjects, but were omitted altogether or grossly confused by the other subjects. These tended first to recall the "long-standing family feud," in such terms as "a common tale of family feud and killing in a primitive society" or as "a Romeo and Juliet story," and this part of the theme was the basis on which the rest of the story was reconstructed. On the other hand, in the recall of those who had seen the picture the main stress fell on the wedding feast, and nearly all of them referred to the picture as a first and dominant association; as one subject said, "the picture created a sort of setting and was the starting-point for the remembering of the story."

Most said that they could remember the picture better than the story. Generally subjects could not decide whether a detail which had been recalled came from the picture or the story, and mistakes were numerous, picture details appearing as intrusions in the recall of the story, e.g. "labourers with coloured hats"; "image of servants carrying plates on a tray"; "sheaves hanging on the wall over a table." These examples illustrate a confusion of picture and story details which was typical.

IV

DISCUSSION

This experiment provided a realistic demonstration of the influence of one experience upon the recall of another. The story was written and presented in a manner intended to give it emphasis and character and to aid its distinction from experiences not under experimental control. In other respects, as has been said, story and picture were deliberately chosen to favour their merging with one another in memory, and several devices in the framing of the instructions to the subjects were employed for this purpose. As would be expected, these intentions were not realised completely. Although this aspect of the results has not been dealt with, the story did not escape the influence of the stream of experiences which preceded and succeeded it, and into which the experiment was inserted, for the recall of the subjects of both groups, A and B, showed much transformation and rationalization from this cause. Again, the picture and story were not entirely assimilated into one another, and each retained some degree of separate identity.

Now it can readily be demonstrated that under conditions, other than those which obtained for our subjects, the two experiences can be perceived, retained, and recalled without contamination of each other. The conditions of our experiment, however, caused the postcards to be viewed with the expectation that at least one of them would be related to the story. As a result six postcards were looked at cursorily and rejected as unrelated; they exerted no detectable influence upon the subsequent recall of the story. "The Village Wedding," however, was recognized as relevant to the story and was examined under the influence of the attitudes induced by it. Being evoked, these attitudes became modified by the picture; and thus modified, they governed the subsequent recall of both picture and story.

The evolution of attitudes in such processes as these may perhaps be conceived of as akin to the shaping of a skilled response by every new situation which evokes it.

If there is indeed a parallelism between the evolution of the attitudes governing perception and recall and the development of a skill, it is not surprising that neither the original attitudes nor the experiences can be recalled unmodified by subsequent experiences. At any rate, neither picture nor story could be recalled in our experiment, even after the shorter intervals, in anything like the forms which they would have taken, had they been independent of one another. No theory of remembering would be satisfactory, therefore, which treated them as independent experiences with a separate identity in memory. In the terms of Bartlett's theory, however, they may be held to have been retained in memory, each as a part of an actively-organized setting, the schemata representing them becoming closely connected to one another and compounded to a degree increasing with passage of time. The progressiveness of the fusion of the experiences argues against any theory which treats the reorganization of memories as similar to the reorganization of a configuration with the addition or subtraction of an element. Evidence for this progressiveness is found in the comparison of recall after less than 8 days with that after 21-28 days, and with that after a year, when it seemed that picture and story had become inextricably mixed with one another.

Comments may now be made upon the nature of the interaction between the two experiences. Undoubtedly the effect of the story on the perception and recall of the picture was due to its verbal character as well as to its precedence in time. "Vocalization," Bartlett states, "favours the classification of presented material according to some rule or *idée directrice*." Our findings showed this tendency to be an important one. "The Village Wedding" was selected because its features complied with the general impressions which the story had created. So far as they can be conveyed in a few words, these impressions were: feud, smouldering enmity, uneasy truce, conviviality, in a rustic, possibly Flemish setting. "The Village Wedding" served as a representative of this general class of scenes, being identified as the correct representative because of the congruence of certain details, such as the crossed sheaves of corn. The story thus provided modes of abstraction, or principles of organization, which caused the features of the picture to be interpreted in terms of concepts of feud, disorderliness, conviviality, and the like. These concepts were not available to the subjects of Group B, who enumerated the particulars of the picture without generalization.

Another aspect of this effect of story on picture is of interest. Whereas the scene portrayed by "The Village Wedding" is usually regarded as one of gaiety, the story suggested it to be one of restlessness and suppressed excitement. Subjects could not rid themselves of this impression, although they might be aware of its origin and falseness. This tendentious attitude was stabilized and perpetuated to some degree by misinterpretations for which it was itself responsible; for instance, innocent and harmless musicians were perceived as servants threatening with weapons, and the throng at the back of the room was taken to be an unruly one. Probably misinterpretations of this kind have counterparts in everyday life, e.g. in rumours and paranoid ideas, when they corroborate or even enhance attitudes which have been adopted as the result of an experience of some special significance. Nevertheless, although sustained to some degree by misinterpretations, the impressions of feud and disorderliness were weakened by the picture, for those who had seen it emphasised conviviality, rather than feud and enmity, in their subsequent recall.

In its turn the picture had thus a considerable effect upon the recall of the story. Not only did it bring about a change in emphasis, but it also facilitated the recall of some features, led to the neglect of others and brought about intrusion of its

details. Possibly some of these results came about the more readily because it favoured the use of visual imagery rather than language; to this tendency may have been due the large number of intrusions. But the most striking effect (although one not easily quantified) was the change in emphasis. This was shown in a marked form in the recall of the story after a year; in the one group the story was reconstructed around the concept of a feud in a peasant society, in the other around the fragmentary recall of a wedding feast.

In conclusion, we may return once more to the question asked in the first section of the paper: what conditions are necessary to bring about the processes of merging of experiences which we have now described? Although it proved no hypothesis, the experiment suggested that the essential condition was the re-arousal by the interpolated material of the memories of the original material and of their attendant attitudes. This being done, a secondary condition was the congruence of some salient details.

APPENDIX

(The story, showing the arbitrary divisions into items)

"For many generations, | the peace | of the little village of Neuf | had been disturbed | by the feud | between the Weydens and the Loons. | Ancient grudges, | constantly revived, | had kept the two families at loggerheads, | and the slightest cause of friction | had been sufficient to fan a smouldering jealousy into a flame of murderous strife. | One day, however, | an aimless sheep | straying from Martin Weyden's field | into that of Henrik Loon | became the cause of a violent quarrel, | in which even the village folk were divided into two camps, | and which ended in the death of Henrik Loon, | the head of the Loon family. | Shortly before his father's death, Henrik Loon's heir | began secretly to court Hetta Weyden, | the daughter of his father's murderer, | and, becoming the head of the family, | hastened to declare his wish to abandon former rivalries and to cement a new cordiality between the two families | by marriage. | Martin Weyden accepted the gesture with generosity and promised a sincere welcome to his new son-in-law. | The village folk were eager to forget their partisanship | and soon showed themselves willing to vie with one another | only in the elaborateness of their preparations for the wedding ceremony. |

The wedding feast, which was held at the Weyden farm, | was honoured even by the sheriff, whose efforts had often been necessary to curb the violence of their quarrels | and to arrange a truce between the families. | Sitting in a high-backed chair | at the head of the broad and laden table, | a dark Holbein cap | almost hiding his white hair | and the velvet of his nobility being somewhat incongruous | in the humble interior of the farmhouse, | he gazed anxiously | at Hans Loon, | the uncle of the bridegroom and the brother of the murdered man. | His solemn expression was unsuited | to the rustic boisterousness of the village folk. | But he had reasons to be alarmed, | for several unhappy incidents | made him doubt the permanence of the sudden friendship between the two families. | The bridegroom | and his relatives | had been greeted with the graciousness and the lavishness which the occasion demanded, | but in a prominent position | over the bridegroom's seat | there hung | the crossed sheaves of corn, | which were not only the traditional symbol of fertility in that agricultural community, | but were also the emblem of the Weyden family. | It could hardly have escaped the notice | of the Loons | that the rival emblem was thus flaunted; | and, on the other side, it was only too obvious | that the bridegroom's coat was made, | as all knew, | from the skin of the sheep | whose straying had led to the death of Henrik Loon. | Upon the bridegroom's head | was

the blue cap | which had been worn last by Henrik Loon on the day of his death. | Indeed, ugly memories | seemed to haunt the young man, | as he sat looking sad and downcast | and hardly interested in the gay chatter of his bride.

"But nothing seemed to daunt the vivacious gossiping of the women, | and as dish after dish was brought | for the guests' pleasure, | and the coloured earthenware pitchers were filled | again and again | with sweet, white wine, | the danger that rancour would flare up again | seemed to grow less. | But just as the men's uneasiness was giving way to a careless enjoyment | of the many delicacies spread before them, | Hans Loon's two servants | entered the farmhouse, | bearing the two ash staves | which they had brought with them to the gathering. | Hans Loon had bidden them stand outside | and guard the horses and saddlebags of his family. | But, seeing the freedom with which the village folk took part in the celebrations, | the two oafish peasants | followed the stream of villagers into the farmhouse. | Although their arms threatened, | their faces showed only a yearning to partake of the rich food | which was being enjoyed by the villagers and the guests alike. | Their countenances were childlike in their stupidity, | as unable to restrain their appetite saliva dribbled from their mouths. | Yet Martin Weyden, angered by their rudeness | and by their display of arms, | called loudly upon Hans Loon | to dismiss them. | But Hans Loon had fallen asleep, | his burly figure slumping forward | and his long white beard falling upon the table. | His sword dangled by his side, | and his faithful hound | peered from under the table | at the strange human ceremony | where the tragedies of enmity were forgotten, | and where fears were lost | in the unrestrained joy of the wedding feast."|

V

REFERENCES

- BARTLETT, F. C. (1932). *Remembering*. Cambridge.
 FREUD, S., with BREUER, J. (1892). In: *Collected Papers*. Vol. I. London.
 WHITFIELD, J. W. (1947). Rank correlation between two variables, one of which is ranked, the other dichotomous. *Biometrika*, **34**, 292-296.

(Received 2nd January, 1950.)

THE RELATIONSHIP BETWEEN AREA AND THRESHOLD IN THE CENTRAL FOVEA FOR LIGHTS OF DIFFERENT COLOURS

BY

E. N. WILLMER

(From the Physiology Laboratory, Cambridge)

The relationship between the size of the field and its threshold intensity has been plotted for small fields fixated on the central fovea. Within the limits investigated, no significant difference has been found whether this relationship has been determined for red or for violet light. While it is possible that there is a continuous relationship between the threshold intensity and the size of the field, in practice there is a rather sharp change in the relationship when the field size subtends about 5' or 6'. Between this size and a field subtending about 20', the expression $\sqrt{A} I = C$ describes the findings, but with smaller fields the relationship $A I = C$ is much more nearly, though not absolutely, true. Experimental difficulties may limit the extent to which the observations can be described by this second equation.

I

INTRODUCTION

It has been known for many years that an inverse relationship exists between the area of the retina stimulated and the brightness of the stimulating light which is necessary to evoke a sensation. For small areas, and particularly in the fovea, Ricco's law is approximately true and states that for threshold excitation the area times the intensity is constant. For larger areas and in the periphery of the retina, Piper's law is more nearly applicable. This law states that, at threshold, the square root of the area times the intensity is constant. Neither of these laws, however, exactly fits the facts, and Wald (1938) has introduced an expression, $(A - n_0)^k I = C$, which satisfied his data and other data for all but very large and very small fields. In this expression, A = area stimulated, n_0 = a constant threshold number of elements, I = intensity and k and C are constants. This author came to the conclusion that two separate laws (Ricco's and Piper's) were not necessary, but that the data could be described by this one equation. Graham, Brown and Mote (1939) showed that any equation of the form $A^k I = C$ was unsatisfactory and produced another quantitative account of the area-intensity relationship which held good in the fovea up to a limit of about 1°, at which point rods probably began to vitiate the results. In the periphery a similar relationship was established which held good for field sizes between the limits of 10' and about 10°, after which the threshold became independent of area. In order to establish whether this peripheral relationship was different for rods and cones Graham and Bartlett (1939) worked with both red and violet light in the periphery. They came to the conclusion that since the curves expressing their results for the area-intensity relationship for the two colours were parallel, there was no evidence that anything but rods contributed to the determination of the threshold. Baumgardt (1947), however, came to somewhat different conclusions. Arguing from van der Velden's experimental finding that two quanta absorbed within a certain time interval are sufficient to produce excitation, he was led to predict that Ricco's law ($A I = C$) should hold for areas of the retina

occupied by one independent or quasi-independent unit, i.e. receptor cell or more probably ganglion cell, and that Piper's law ($\sqrt{AI} = C$) should hold for areas containing more than one such unit. His experiments initiated by this hypothesis showed that with a blue light at a position 15° temporally from the fovea, the relationship $AI = C$ held for rod vision for areas between $2' 12''$ to $31' 36''$, and thus the quasi-independent unit appeared to be large and presumably might correspond to the ganglion cell which is known to spread its dendrites far and wide and to cover a large number of rods. On the other hand, when working with extreme red light and finding no photochromatic interval, so that he was probably stimulating cones, he found that Piper's law ($\sqrt{AI} = C$) was obeyed between $12' 30''$ and $31' 36''$, indicating the presence of many quasi-independent units within this area and suggesting that the ganglion cells activated by cones occupy smaller areas, which is of course in accordance with the histological findings (Polyak, 1941). More recently (Baumgardt, 1948) he has extended the observations to the fovea and has shown that under several different conditions with respect to duration of the flash and to pupil size there was no significant difference between the behaviour of the fovea in red light or green light. For example, he found that when no artificial pupil was used, the relationship $A^{0.95} I = C$ was obeyed within the limits $1.12'$ and $6.72'$ diameter of field. Between $6.72'$ and $33.6'$ the relationship was $A^{0.42} I = C$, thus showing that neither Ricco's nor Piper's law was exactly valid but they are each nearly so for the field size to which they are appropriate and that there is a change of law at about $6.72'$.

Now, if Baumgardt is correct in supposing that the range within which Ricco's law holds good gives a measure of the size of the quasi-independent unit, then the foveal unit must correspond in size to the retinal field covered by an object subtending an angle of $6.72'$. This is considerably above the size of field corresponding to a single cone but might relate again to the area served by a single ganglion cell or by certain types of bipolar cell, although the data for the visual acuity of the fovea would indicate the probability that the effective unit is considerably smaller than this. If, however, the size of the quasi-independent unit is the determining factor in bringing about the change of law, then it is relevant to enquire whether this unit differs according to the wavelength in a manner which might suggest different pathways for different colours. Since the central fovea behaves as though it had only two independent types of receptor and two corresponding pathways (König, 1894; Willmer, 1944) it is probable that one of these will be stimulated more intensely by light from the blue end of the spectrum while the other will respond more in the red region. It therefore seemed possible that some information might be obtained as to the size of the quasi-independent unit for each of these pathways by making measurements of the area-intensity relationship for the central fovea in red light and in violet light. As already mentioned, Baumgardt (1948) has found that there was no significant difference when red and green lights were used, but it might be that these two wavelengths are not far enough apart to stimulate the two receptor systems sufficiently exclusively to make any significant difference. This consideration together with the fact that there was a divergence between Baumgardt's observations with red and blue light in the periphery and the earlier observations of Graham and Bartlett, made it desirable to re-investigate the area-intensity relationship in the central fovea and to use violet and red test lights instead of green and red. By this means it might be possible to determine the size of the quasi-independent units in the fovea and to decide whether or not the units corresponding to the two receptors differed significantly in size.

II

METHODS

The apparatus already described (Willmer, 1949) was used to illuminate two small fields with light of any colour or intensity required. The right-hand field subtended about $5'$ at the eye and was illuminated throughout the experiments by light from an Ilford spectrum red filter (608) at an intensity about five times the threshold for the foveal centre of the subject. This field acted as a pilot light or fixation point and the subject was required to fixate on the left-hand edge of this field. By means of stops, the test field, which could be illuminated by either extreme red (Ilford 609) or violet (Ilford 601) was located with its centre $20'$ away from the centre of the pilot field and horizontally to the left. The stops varied in size in such a way that, at the eye, their diameters subtended angles varying between $0.72'$ and $19.4'$. With accurate fixation these fields would all fall on the so-called rod-free area. Fixation was assisted by the usual dental impression technique for maintaining a constant head position. An artificial pupil, 2 mm. in diameter, was used throughout the experiments. The subject could vary the intensity of the light in the left-hand field by turning a knob and he was allowed as much time as he liked to reach a decision as to what was the threshold position when he was certain of his central fixation. In this way the technique differed to some extent from that adopted by previous investigators. Between each reading the position of the lamp on the scale was altered, thus changing the light intensity. This was done by the operator so that the subject could have no idea from the position of the lamp or of any other part of the apparatus whether he was near threshold or not and the complete independence of each reading was thus secured. The fields of different sizes were given in random order, and the readings were repeated on several different occasions, so that any point in the diagrams is always based on the mean of at least four entirely independent readings and most of them are based on ten or more. The readings have all been made on two subjects whose vision and colour-vision were normal.

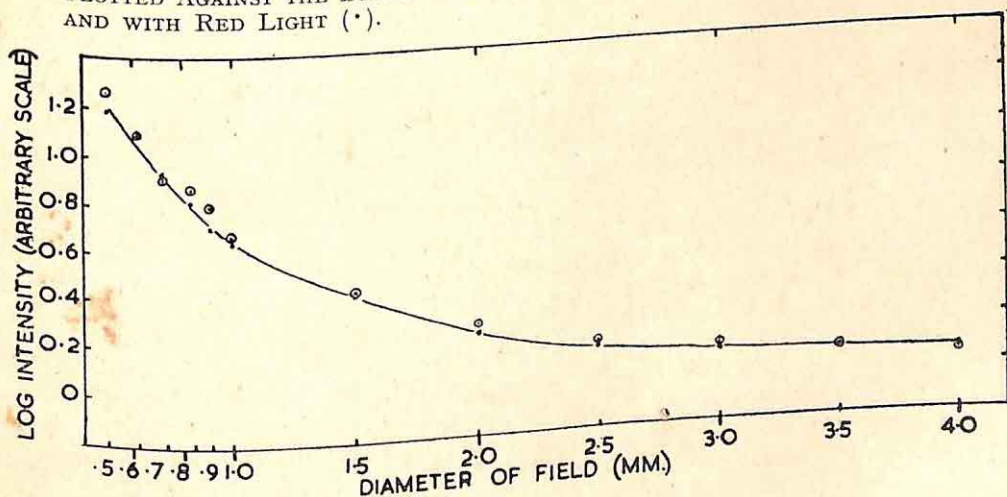
III

RESULTS

The first experiments were designed to test whether violet or red test fields made any difference to the relationship between intensity and field-size. The threshold intensities were recorded with central fixation for each size of field from $2.4'$ to $19.4'$, and in order to eliminate day-to-day fluctuations in absolute threshold the logarithms

FIGURE 1

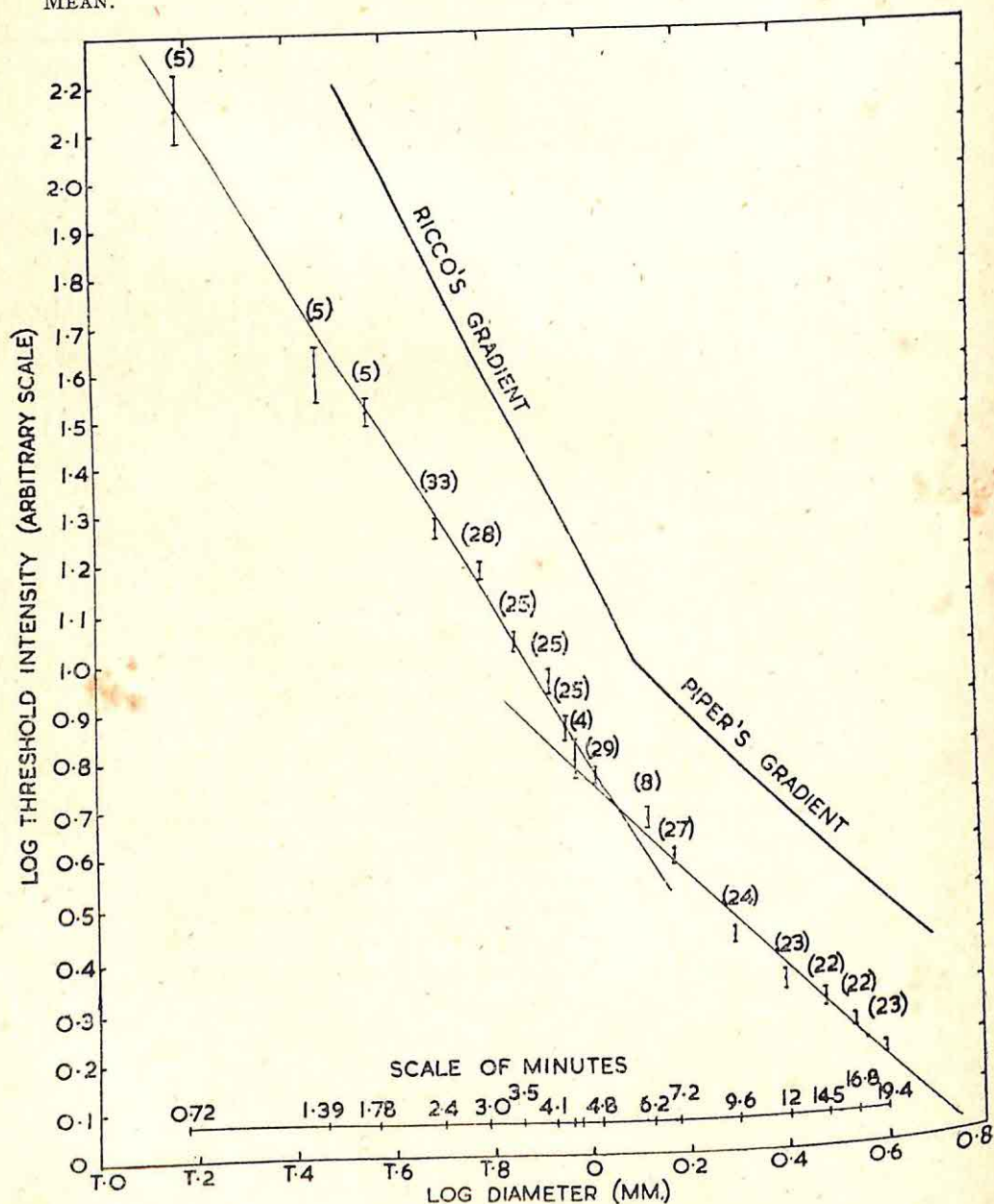
THE LOGARITHM OF THE INTENSITY OF THE THRESHOLD STIMULUS IN ARBITRARY UNITS, PLOTTED AGAINST THE DIAMETER OF THE FIELD IN MM., WITH VIOLET LIGHT (\odot) AND WITH RED LIGHT (\cdot).



of the intensities were plotted against the diameter of the field and the readings for each day were moved along the vertical axis in such a way as to obtain the best fits with preceding readings. The results are recorded in Figure 1. The intensity values are in arbitrary units. From this figure it is clear that very much the same relationship holds good for both the violet and the red test fields. The range of the

FIGURE 2

THE LOGARITHM OF THE THRESHOLD INTENSITY PLOTTED AGAINST THE LOGARITHM OF THE DIAMETER OF THE TEST FIELD IN MM. THE MEAN OF THREE SERIES OF READINGS ON TWO OBSERVERS; THE NUMBERS IN BRACKETS INDICATE THE NUMBER OF READINGS. THE VERTICAL LINES INDICATE THE STANDARD DEVIATION OF THE MEAN.



individual points upon which the points plotted in Figure 1 are based is of the order of 0.2 log units, so that the position of the circles (violet light) is nowhere significantly different from that of the points (red light). If Baumgardt is right in his hypothesis with regard to the interpretation of Ricco's and Piper's laws then it is clear that whatever relationships are established for the "red" receptors are equally valid for the "violet" receptors so long as vision is confined to the central fovea; it is therefore impossible to separate the two types of receptor by their behaviour with regard to threshold intensity and area stimulated, at least with field sizes down to 2.4' in diameter.

TABLE I

<i>Log diameter of field</i>	<i>Angle subtended by field in minutes</i>	<i>Log threshold intensity, mean values (numbers of observations in brackets)</i>	<i>Standard deviation of the mean</i>
1.18	0.72	2.15 (5)	0.07
1.46	1.39	1.60 (5)	0.06
1.57	1.78	1.52 (5)	0.03
1.70	2.4	1.27 (33)	0.02
1.79	3.0	1.18 (28)	0.019
1.86	3.5	1.04 (25)	0.016
1.93	4.1	0.96 (25)	0.024
1.96	4.4	0.87 (25)	0.020
1.98	4.6	0.80 (4)	0.04
0.02	4.8	0.77 (29)	0.014
0.13	6.2	0.68 (8)	0.020
0.18	7.2	0.60 (27)	0.016
0.30	9.6	0.43 (24)	0.014
0.40	12.0	0.34 (23)	0.022
0.48	14.5	0.30 (22)	0.016
0.54	16.8	0.25 (22)	0.013
0.60	19.4	0.19 (23)	0.010

The further question, however, remained whether the results are more adequately described in terms of some continuous relationship as suggested by Wald (1938) and by Graham, Brown and Mote (1939), or whether there is in fact a change from something akin to Ricco's law when the fields are very small to something more resembling Piper's law when the fields exceed a certain critical size, this size then representing the size of the quasi-independent unit as suggested by Baumgardt. Figure 2 represents the mean of three sets of results relating the logarithm of the intensity of the stimulus to the logarithm of the diameter of the field in mm. The figures (see Table I) were obtained from two normal observers, one set for violet light and the other two sets for red light. There is, in spite of the fact that most of the points are based on the means of more than 20 readings, some considerable margin of error in the exact location of the points and it would be possible to draw a continuous curve through the points as plotted. Nevertheless, there is a rather strong suggestion that the two straight lines drawn in the diagram might fit the points almost equally well. If this suggestion is accepted it would mean that one relationship holds for fields smaller than about 5' or 6' (log. 0.06 mm.) and another for fields larger than this. The other two lines on Figure 2 define the gradients

corresponding to Ricco's and Piper's laws, and clearly show that, between 5' and 20', Piper's law is very nearly true. With diameters less than 5', however, there is a very considerable divergence from Piper's law and an approach to Ricco's law, though the agreement is not so satisfactory.

The position therefore seems to be that, while the data may be capable of description by some continuous function, in effect, a rather pronounced inflexion occurs at a field size which subtends about 5' or 6' at the eye. Above this size, the relationship $\sqrt{A} I = C$ is substantially true, while below this size the relationship approximates more nearly to $A I = C$.

IV

DISCUSSION

The results of this analysis thus, in some respects, confirm and extend those of Baumgardt (1947, 1948) and if he is right in supposing that the change from Ricco's law to Piper's law occurs as soon as more than one independent or quasi-independent unit is involved, then it would appear that the area occupied by such a unit in the central fovea corresponds to a field subtending about 5' or 6' at the eye. This is manifestly too large to correspond to a single receptor for which the figure would probably be more of the order of 30". If therefore it has any anatomical counterpart, this is more likely to be the flat-topped bipolar cell or the ganglion cell. The dendritic tree of the flat-topped bipolar cell is of the right order of magnitude and corresponds to a field size of about 6' though it is probably somewhat smaller than this in the foveal region. If the size of this anatomical unit were shown to be connected with the change of law, as these results perhaps suggest, it is pertinent to ask whether any similar change in the area-intensity relationship occurs with still smaller fields corresponding in size to the midget bipolar cells. Is the divergence from Ricco's law with the smaller fields seen in Figure 2 perhaps explicable on the grounds that both midget bipolars and flat bipolars are involved, and would the law become rigidly true at a field size corresponding to the midget bipolar cell? The present methods and results do not give the information necessary to answer such questions. In the first place, very few figures have been obtained for the three smallest field sizes and these with red light only, on account of the practical difficulty of obtaining a sufficiently intense source of violet light. Secondly, it is clear that very small errors in the size of the stops delimiting the fields, when these are of the order of 0.5 mm. in diameter have a very significant influence on the threshold values obtained for these small fields. Moreover, and this is probably the most important point of all, the optical perfection of the image on the retina becomes questionable at these small dimensions and the actual form of the image on the retina is a matter of theory and guesswork. It is clear therefore that these suggestions as to the correspondence between the figures for the smallest fields and the anatomical units concerned are little more than ideas for future research. The only justifiable conclusion from the present data is that with fields smaller than about 5' or 6' in diameter, the product of area times intensity is much more nearly constant than that involving the square root of the area. Whether the relationship becomes rigidly true for areas subtending 1' or so, cannot be determined from the data available. The fact that the square root relationship is so nearly true for fields subtending between 5' and 20' and then diverges so widely with smaller fields does perhaps indicate a critical point corresponding to fields of about 5', and this would support the idea that some receptive unit in the central fovea has approximately these dimensions. It is clear, however, that the figures here presented for

the area-density relationship in the central fovea allow no decision to be reached as to whether in fact this apparent break at about 5' is anything more than an inflection in a continuous curve relationship.

The author is indebted to the Leverhulme trustees for a grant enabling this investigation to be completed. He also wishes to express his gratitude to the two subjects (D. G. and P.M.H.R.) for submitting to the ordeal, and to W. A. H. Rushton for much constructive criticism.

V

REFERENCES

- BAUMGARDT, E. L. M. (1947). *J. gen. Physiol.*, **31**, 268.
BAUMGARDT, E. L. M. (1948). *C.R. Soc. Biol., Paris*, **142**, 464.
GRAHAM, C. H. and BARTLETT, N. R. (1939). *J. exp. Psych.*, **24**, 574.
GRAHAM, C. H., BROWN, R. H. and MOTE, F. A. (1939). *J. exp. Psych.*, **24**, 555.
KONIG, A. (1894). *S.B. Akad. Wiss., Berlin*, 577.
POLYAK, S. L. (1941). *The Retina*. Chicago.
WALD, G. (1938). *J. gen. Physiol.*, **21**, 269.
WILLMER, E. N. (1944). *Nature, London*, **153**, 774.
WILLMER, E. N. (1949). *J. Physiol.*, **110**, 377.

(Received 12th December, 1949.)

AN EXPERIMENTAL STUDY ON THE FITNESS OF SIGNS TO WORDS

BY

K. R. L. HALL and R. C. OLDFIELD

(From the Institute of Experimental Psychology, Oxford)

This experiment, following Bartlett's Picture-writing, was designed to show: how much agreement there would be in the choice of a sign as the most fitting in connection with a word; whether the reasons given for a choice of sign would vary greatly, and would be rationalizations rather than a primary ground for so choosing; and whether the chosen signs were more easily recalled than the not-chosen.

Six cards were shown to the subject, on each of which was one word and two signs. The signs were not intended to represent any object, although none was arbitrary. The subject had to choose the more appropriate sign and give the reason for his choice. Afterwards he had to reproduce all the signs.

Agreement in choosing one of the signs of each pair was surprisingly high, although the reasons varied considerably. The chosen signs were reproduced more accurately and frequently, the reproductions showing the conventionalizing effect of the reasons given.

Choice of a sign as most fitting seemed to depend upon certain formal structural qualities of a general character. The most fitting sign for a verbal setting seemed to combine a conventional structure with simplicity and regularity of design. Linking these results with those of an experiment of Harrower's, it is considered possible that a method may here be available for exploring forms of mental organization not so far tapped by standard intelligence tests.

I

INTRODUCTION

In his Picture Writing experiment, Bartlett (1932) used a variety of signs which were more or less arbitrarily selected or constructed, and the subjects were required to use these signs in place of the words with which they were associated. Some of the signs were directly representative, some were designed to awaken secondary associations and some were "of the type commonly, but inaccurately, called 'meaningless'" (p. 97). As he pointed out, however, obviously no single classification could be made owing to the fact that every subject would classify such material in his own way. The general task was to learn the signs in connection with the words accompanying them, and, when the words were read out later on, to reproduce the signs they had learned for them.

In the "effort after meaning" required by the task and the nature of the material, a variety of reactions were observed in the subjects. The representative signs were "easy," and the connection obvious, but where a subject was unable to give a sign any precise and unambiguous representative significance, it would almost certainly be omitted in the reproductions.

Where a common and conventional setting was not provided by the word with which the sign was connected, an effort was frequently made to find some alternate form of setting. This might be achieved by finding some familiar secondary association, or some particular connection which could not be easily formulated in words but which was felt to be appropriate "in that peculiar and unmistakable way which always occurs when some presented material seems to be singularly in

its right place but we do not know why it should be so" (p. 113). For example, for the sign associated with the word "thing," one subject appeared to "feel" its particular suitability and pleasingness, and the sign for "Philosophy" was reported by several subjects to be peculiarly fitting.

Elaborating on this affective fittingness, Bartlett observed (p. 114) that the satisfaction which accompanied such an experience could occur both when the subject was able to say why the sign fitted, and when he was unable to formulate any reason. Two of his subjects had, for instance, found this same "Philosophy" sign particularly pleasing "simply because it was a big structure tottering on a little foundation." He concluded that such fitness of certain material may involve complex cognitive processes which do not need to be formulated in order to be appreciated.

Now these observations of Bartlett's were more or less incidental to the main purpose of his experiment, and it seemed that there might very well be important individual differences identifiable here in the reactions to similar sorts of material. The less "familiar" and "conventional" the material, the more probable it would be that such subjective differences might be observed.

We can therefore summarize certain of the implications of these findings as follows:

- (1) The fitness of a sign in connection with a word need not be on account of any *representational* appropriateness or secondary associative significance.
- (2) The peculiar fitness of a sign may be manifested by an affective reaction on the part of the subject. Any associative significance that may afterwards be formulated may therefore not necessarily have been operative in this primary "feeling" reaction. The formulation may therefore be in the nature of a rationalization.
- (3) It is possible that individual differences in the way in which subjects deal with such sign-word combinations may be important where the connections are not conventional and easy.

The present experiment was therefore designed to show:

- (1) How much agreement there would be, in a group of subjects of similar background, in the choice of a sign as the most fitting in connection with a word.
- (2) Whether the reasons given for choosing a particular sign would show much diversity, and would seem to reflect more a subsequent rationalization than a primary ground for the choice.
- (3) Whether the chosen signs were more easily and accurately remembered than the not-chosen.

II

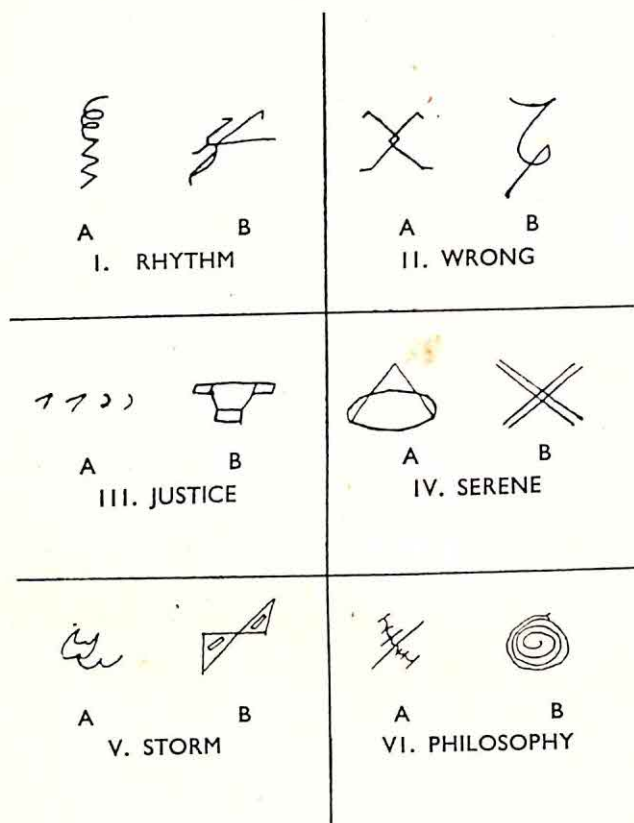
DESCRIPTION OF THE EXPERIMENT

The Method of Choice was used, that is, a pair of symbols was presented in connection with a word, and the subject's task was to choose which of the two he thought to be the more appropriate.

The materials used consisted of six plain white cards, 3 by 5 in., upon which were drawn in black ink two symbols or signs, under which was printed a word. The set of twelve signs and the six words which accompanied them are shown below. Two of the signs were the same as those used by Bartlett in his Picture-writing experiment, sign A for Philosophy, and sign B for Wrong. In drawing the remaining ten signs, it was intended that they should be similar to those that Bartlett had used in his third set, and direct object-association was avoided as far as possible. They were, however, drawn

in connection with the words they accompany, that is, they were not arbitrary, and not randomly selected, but there was no intention of making one or other of a pair the more significant member.

FIGURE 1



The subjects who took part in the experiment were all of university educational standard, both male and female, and represented a fairly wide range of intellectual background.

There were three main parts to this experiment, in each of which the experimental conditions were slightly varied. The essential task was the same, that of choosing a sign with reference to a word, throughout the experiment.

Preliminary. 17 subjects—8 Male, 9 Female.

This was mainly to help in deciding the best experimental procedure. Each of the six cards was presented to the subject for 10 seconds, the order in which they were presented being altered from subject to subject. The subject was instructed to choose the A or the B symbol, according to which he thought was the most fitting or meaningful, but might say "neither" or "both," if he was unable to judge between them. He was asked to state the reason for his choice.

Fourteen of the 17 subjects were asked, immediately after they had finished their choices, to reproduce both the symbols which they had been shown in connection with a particular word. The six words were printed on a sheet of paper, and the reproductions were made in an A and B column opposite. There was no time limit to this part of the experiment, and no previous indication had been given to the subjects that they would be required to recall what they had seen on the six cards.

Main Experiment. 40 subjects—29 Male, 11 Female.

- (1) Using the same material, the procedure was modified. The subjects were instructed definitely to choose either A or B, and no time limit on the making of the choice was imposed. Where a subject could not decide between A and B,

he was eventually asked to guess, as this was found invariably to elicit a choice, usually backed by a reason.

- (2) One variation was introduced, in order to determine whether the position of the signs on the card might be a factor in biasing the choice, and, for 10 of the subjects, a second set of cards was used, with the B sign of the original set on the left and marked A, so that the position was reversed. The procedure was identical.
- (3) Ten subjects were asked for reproduction of the symbols, as in the Preliminary Group, except that here there was an interval of 10 minutes between making the choices and the reproductions. Seven other subjects were also asked to reproduce the symbols, the delay being varied from five minutes up to three weeks. For two other subjects, the reproduction procedure was reversed, and, instead of being given the words, the 12 signs were shown singly, in random order, and the subject was required to respond with the correct word for each.

Group III. 20 subjects—18 Male, 2 Female.

In order to find out whether certain of the symbols were predominantly chosen irrespective of the verbal context in which they were framed, that is, whether, in at least some cases, the relationship between word and symbol was much less specific than might have been supposed, the same pairs of symbols were shown, but with the words blanked out so that they were not in any way visible. The subject was instructed as follows: "There are six cards here, each with two symbols drawn on them. Before I show them to you, I am going to tell you a word, and, when I show you the first card, I want you to consider the two symbols on it in relation to that word, and choose which of them is most fitting and meaningful in that connection. With the same word in mind, I want you to choose one of the two symbols on each of the other five cards."

When all the six cards had been shown, with reference to the first word, they were spread out on the table in front of the subject, and he was asked to say which symbol he thought was the best in connection with that word.

The same procedure was carried out until all the six words had been used, and choices from every pair of symbols made in connection with each. The order of the words was varied from subject to subject, as was the order of presenting the symbols in connection with the words.

The subject was not asked to give any reason for his choice.

III

RESULTS

Distribution of Choices

(1) *General Distribution.*

For each of the six cards, a very substantial measure of agreement of choice was found. Table I shows the actual distribution of choices made, for the 40 subjects of the main experiment.

TABLE I

<i>Card</i>	<i>Number choosing A</i>	<i>Number choosing B</i>	<i>P</i>
1	34	6	≤ 0.01
2	31	9	≤ 0.01
3	8	32	≤ 0.01
4	31	9	≤ 0.01
5	33	7	≤ 0.01
6	3	37	≤ 0.01

We may examine this distribution relative to the hypothesis that neither figure on each card is more likely to be chosen than the other. Column P in Table I shows the probabilities of the observed distributions, or their opposites (viz. (40-A), (40-B),

every case clear which figure was intended. Only the *correct* reproductions (R) are counted in Table VII, and those which are clearly incorrect and which show no recognizable similarity to the originals are classed with the not-reproduced (NR).

The difference between the number of chosen and the number of non-chosen figures reproduced is highly significant for both groups of subjects, while there is no significant difference obtained from comparing the Preliminary Group frequencies with the Main Group, the value of chi square being 0.0024, which, for one degree of freedom, gives a probability far higher than 0.05. Thus, approximately twice as many of the figures chosen by the subjects as the more appropriate are reproduced with reasonable accuracy.

Regarding the actual number of reproductions made for each figure, these are shown in relation to the number of choices made for that figure in Table VIII. For the Preliminary Group, where the judgment was "neither" or "both" in choosing for a card, this has been omitted from the number of choices shown, only definite A and B choices being counted. Hence the totals do not come to 14.

TABLE VIII
Preliminary Group

	I		II		III		IV		V		VI	
	A	B	A	B	A	B	A	B	A	B	A	B
Chosen	8		11	1	2	10	6		8	3	2	9
Reproduced	10	8	11	6	7	13	10	12	13	10	5	14

Main Group

Chosen	9	1	9	1	2	8	8	2	7	3	0	10
Reproduced	10	4	10	3	5	9	7	7	7	3	2	10

There is clearly a close relationship between Choice and Reproduction, particularly with the Main Group, where judgment was required to be a definite A or B. For this Group, only for Card IV is the number of reproductions equal for each figure, and this is probably due to the fact that some subjects gave a negative reason for choosing A, by saying that B is cross, or a busy cross-roads, and therefore is the opposite of Serene. The verbalization which the subject made very often influenced the reproduction of the figure to which it referred. It will be observed that Figures I.A, II.A and VI.B are particularly privileged both in Choice and Reproduction for this Group.

Qualitative Changes in the Reproductions

(I) *Conventionalization.*

Many of the changes in the reproductions, when compared with the original figures, were clearly due to Object Assimilation or Conventionalization. The subject was required to give his reason for making a choice, although no attempt was made to influence him if he was unable to formulate the reason, and some subjects rested content with giving no particular reason or only a very vague statement of it.

Considered in conjunction with these verbal reasons, the effect of the verbalization upon the reproductions is obvious, and, in some cases, quite striking. Figures such as I.A for Rhythm, II.A for Wrong, and VI.B for Philosophy, apparently lend

themselves very readily for easy assimilation and reasonably accurate reproduction. Certain small details are omitted, and the II.A figure might become simplified to a conventional cross-out sign, I.A might be perceived and recalled as a "regular, smooth-flowing spiral," or as a coiled spring, so that the jagged, lower part of the figure is reproduced as circular, in conformity with the upper part.

A greater variety of assimilative changes can be seen in the reproductions of III.A and V.A, in keeping with the variety of object association that these figures aroused. Hence the reproductions are on the whole less accurate than those for I.A or VI.B, for example. If the same figure is chosen either because it is like a judge's wig, a gallows, a court bench, or scales of justice, the reproductions may be expected to show corresponding deviations.

There is, therefore, no doubt whatever that the verbalization intervening as a reaction between the making of the choice and the making of the reproduction has in many cases influenced this reproduction. The name given, whether it be an object-name such as "a man dancing" or "an empty cross-roads," or a more general name such as "spiral," "circularity," or "zig-zag," is often that which is aroused when reproduction is required. The conventionalizing effect of the name, which Bartlett, Wulf and Gibson observed, is confirmed, while there is further evidence of the simple and symmetrical characteristics of these conventional, conceptual patterns to which the figures are related.

It seems, therefore, that the factors which influence the original judgment of fitness in the choice of a figure in connection with a verbal setting serve to enhance the recall-value of that figure. Simplicity and economy of structure, within the framework of the verbal setting, are both preferred in perception and achieved in reproduction.

(2) *Figure Assimilation.*

This type of change that both Bartlett and Gibson observed was also found in a few cases. So, in three reproductions made of Figure I.B, there is evidence of confusion with Figure II.A.

IV

DISCUSSION

Where, as in the material used in the present experiment, one is dealing with more or less complex relationships between words and symbols, the nature of the real basis of choice does not reveal itself very clearly. The high number of choices falling on one particular sign rather than the other is not supported by a corresponding agreement in the reason for making that choice. What then is the nature of this fittingness of word and sign?

Apart from Bartlett's own experiments, little work seems to have been done on this problem. Harrower (1933), however, carried out an interesting experiment when studying "Organization in the Higher Mental Processes." She thought that "the joke allowed of some sort of diagrammatic representation of the particular relations holding within its parts: directly analogous to the difference in form and structure in the fields of perception" (p. 58). She accordingly presented a set of four diagrams with each joke, three of which were arbitrary, that is they were not intended to bear any relation to the joke, while the fourth was intended to express the relational *structure* of the joke. She found that, not only were most of her subjects able to select the correct diagram of the four, but that the recall value of the jokes when accompanied by the proper diagram was far higher than where the same jokes were accompanied by arbitrary diagrams, and far higher, also, than where

Inasmuch as, in the result for Group I and II, there is a clear and significant consensus of choice, we may speak of the figure chosen less frequently as the "unpopular" figure, and consider the distribution of unpopular responses among the subjects. Table VI shows this distribution.

Among the male subjects, in fact, there were 14 out of 29, or nearly 50 per cent., who gave no unpopular responses at all.

Reasons for Choice.

All the subjects in this experiment were asked, when each choice had been made, to give the reason for their decision. In a few cases (22 out of 240 for Groups I and II) they were quite unable to do so. In the remainder they were able to give *some* reason, although often with a good deal of hesitation, so that one gained the impression that the choice itself was a direct schematically determined reaction like the act of recognition, while the evocation of the reason was less a statement of the causes for the choice than a *justification* of an attitude already evoked in the choice.

One figure might be chosen rather than another, because:

- (1) of some alleged similarity of quality or form between figure and idea (such as circularity for philosophy),
- (2) it suggested some form or object connected with the idea and, in this case, the object or form might be either—
 - (i) one conventionally regarded as a symbol for the idea in question (scales or balance for justice), or
 - (ii) one in some other way connected with the idea (judge's wig for justice, cloud for storm),

and here the association is similar to the conventional object except that it is *actually* to be observed in this connection, whereas the scales of justice are only so shown in pictures or on statues, and so on.

A very small number (7 in 240) of cases were found in which a figure was chosen because it suggested some *activity* connected with the idea (e.g. dancing, for rhythm). And, apart from those cases in which no reason could be given, a few (20 in 240) were unclassifiable under any of the above heads.

Summarizing some of the main points from the reasons given for the choices, it was found that:

- (1) It is some quality, whether of the figure as it stands, unnamed, or whether, of a named, associated object, such as scales, that is the commonest apparent factor in determining the choice. It seems probable, for instance, that it is the *balance* of the popular figure for Justice, rather than any realistic resemblance to a pair of scales, which is important for the choice; that, in fact, it may be a general quality of the figure rather than any specifiable associated object that is influential, for, as is seen in more than one instance, object-associations can perfectly easily be found for the unpopular choices, and it is not, therefore, a matter of objective resemblance.
- (2) The quality of the figure, in terms of symmetry, balance, regularity, and so on, may, where the word accompanying may suggest it, actually serve as a ground for *not* choosing that figure.
- (3) In a few instances, subjects tend to avoid the obvious association, as in the cross-out sign for Wrong.

- (4) There is, particularly for card 5A, a preference for the less stable, and more "moving" figure in connection with storm: this, again, is a perception of a general quality that seems to count most rather than perception of any particular object-likeness—for the very variety of objects evoked is sufficient evidence of the generality of the figure.
- (5) There is no very overt indication of affective reaction, such as Bartlett found, yet the general impression derived from observing the subjects' reactions and from some of the verbalizations regarding the quality and symmetry of a particular figure is that the fittingness of, say, the B figure for Philosophy or the A for Serene is much more *felt* than logically conceived.

Conclusions on the Main Experiment.

(1) In any particular choice set up there may be a very high degree of agreement between different people of the same kind of group about the relative fitness of one or the other figure.

(2) The grounds given for the choice may be very various in their wording, even where there is considerable agreement on choice, and even in the type of reason given, although there may be a common and more general ground for choosing a particular figure discernible through this variety.

Group III.

As a final variation in the choice procedure, arising out of the results just considered, a further group of 20 subjects was shown the six cards with the words blanked out, the procedure being as described on page 63. It was thought that one sign might be chosen rather than another, because it is, in itself and irrespective of its verbal context, especially pleasing, and also that one sign might be considered particularly suitable in connection with several words. For reasons of space the detailed results are not shown, but, from the results it seems that the general impression of the main experiment is confirmed, namely, that it is some formal, qualitative aspect of the sign that makes it either chosen or rejected as the case may be, rather than any object-association being the *primary* ground for choice. A sign is not absolutely fitting in connection with any verbal setting, but is selected in its comparative relation to the other member of the pair, making up the total problem-unit.

Reproduction of the Figures.

Fourteen subjects of the Preliminary Group made reproductions of the figures immediately after they had completed the choice part of the experiment, and 10 subjects of the Main Experimental Group made reproductions after an interval of 10 minutes. As has already been emphasized, the experimental task was specifically stated to be the choosing of the best figure, and no expectation of having to recall the figures was aroused.

TABLE VII

Subject group	Number of subjects	Chosen		Non-chosen		Value of chi square	P
		R.	N.R.	R.	N.R.		
Preliminary	14	79	5	42	42	40.43	≤ 0.01
Main	10	56	4	20	40	46.50	≤ 0.01

Table VII gives the number of chosen figures that were reproduced and the number of non-chosen figures reproduced for these two groups. Although there was considerable variation from the originals in the reproductions made, it was in almost

every case clear which figure was intended. Only the *correct* reproductions (R) are counted in Table VII, and those which are clearly incorrect and which show no recognizable similarity to the originals are classed with the not-reproduced (NR).

The difference between the number of chosen and the number of non-chosen figures reproduced is highly significant for both groups of subjects, while there is no significant difference obtained from comparing the Preliminary Group frequencies with the Main Group, the value of chi square being 0.0024, which, for one degree of freedom, gives a probability far higher than 0.05. Thus, approximately twice as many of the figures chosen by the subjects as the more appropriate are reproduced with reasonable accuracy.

Regarding the actual number of reproductions made for each figure, these are shown in relation to the number of choices made for that figure in Table VIII. For the Preliminary Group, where the judgment was "neither" or "both" in choosing for a card, this has been omitted from the number of choices shown, only definite A and B choices being counted. Hence the totals do not come to 14.

TABLE VIII
Preliminary Group

	I		II		III		IV		V		VI	
	A	B	A	B	A	B	A	B	A	B	A	B
Chosen	8	5	11	1	2	10	6	5	8	3	2	9
Reproduced	10	8	11	6	7	13	10	12	13	10	5	14

Main Group

	A	B	A	B	A	B	A	B	A	B	A	B
Chosen	9	1	9	1	2	8	8	2	7	3	0	10
Reproduced	10	4	10	3	5	9	7	7	7	3	2	10

There is clearly a close relationship between Choice and Reproduction, particularly with the Main Group, where judgment was required to be a definite A or B. For this Group, only for Card IV is the number of reproductions equal for each figure, and this is probably due to the fact that some subjects gave a negative reason for choosing A, by saying that B is cross, or a busy cross-roads, and therefore is the opposite of Serene. The verbalization which the subject made very often influenced the reproduction of the figure to which it referred. It will be observed that Figures I.A, II.A and VI.B are particularly privileged both in Choice and Reproduction for this Group.

Qualitative Changes in the Reproductions

(I) Conventionalization.

Many of the changes in the reproductions, when compared with the original figures, were clearly due to Object Assimilation or Conventionalization. The subject was required to give his reason for making a choice, although no attempt was made to influence him if he was unable to formulate the reason, and some subjects rested content with giving no particular reason or only a very vague statement of it.

Considered in conjunction with these verbal reasons, the effect of the verbalization upon the reproductions is obvious, and, in some cases, quite striking. Figures such as I.A for Rhythm, II.A for Wrong, and VI.B for Philosophy, apparently lend

themselves very readily for easy assimilation and reasonably accurate reproduction. Certain small details are omitted, and the II.A figure might become simplified to a conventional cross-out sign, I.A might be perceived and recalled as a "regular, smooth-flowing spiral," or as a coiled spring, so that the jagged, lower part of the figure is reproduced as circular, in conformity with the upper part.

A greater variety of assimilative changes can be seen in the reproductions of III.A and V.A, in keeping with the variety of object association that these figures aroused. Hence the reproductions are on the whole less accurate than those for I.A or VI.B, for example. If the same figure is chosen either because it is like a judge's wig, a gallows, a court bench, or scales of justice, the reproductions may be expected to show corresponding deviations.

There is, therefore, no doubt whatever that the verbalization intervening as a reaction between the making of the choice and the making of the reproduction has in many cases influenced this reproduction. The name given, whether it be an object-name such as "a man dancing" or "an empty cross-roads," or a more general name such as "spiral," "circularity," or "zig-zag," is often that which is aroused when reproduction is required. The conventionalizing effect of the name, which Bartlett, Wulf and Gibson observed, is confirmed, while there is further evidence of the simple and symmetrical characteristics of these conventional, conceptual patterns to which the figures are related.

It seems, therefore, that the factors which influence the original judgment of fitness in the choice of a figure in connection with a verbal setting serve to enhance the recall-value of that figure. Simplicity and economy of structure, within the framework of the verbal setting, are both preferred in perception and achieved in reproduction.

(2) *Figure Assimilation.*

This type of change that both Bartlett and Gibson observed was also found in a few cases. So, in three reproductions made of Figure I.B, there is evidence of confusion with Figure II.A.

IV

DISCUSSION

Where, as in the material used in the present experiment, one is dealing with more or less complex relationships between words and symbols, the nature of the real basis of choice does not reveal itself very clearly. The high number of choices falling on one particular sign rather than the other is not supported by a corresponding agreement in the reason for making that choice. What then is the nature of this fittingness of word and sign?

Apart from Bartlett's own experiments, little work seems to have been done on this problem. Harrower (1933), however, carried out an interesting experiment when studying "Organization in the Higher Mental Processes." She thought that "the joke allowed of some sort of diagrammatic representation of the particular relations holding within its parts: directly analogous to the difference in form and structure in the fields of perception" (p. 58). She accordingly presented a set of four diagrams with each joke, three of which were arbitrary, that is they were not intended to bear any relation to the joke, while the fourth was intended to express the relational structure of the joke. She found that, not only were most of her subjects able to select the correct diagram of the four, but that the recall value of the jokes when accompanied by the proper diagram was far higher than where the same jokes were accompanied by arbitrary diagrams, and far higher, also, than where

the joke was presented unaccompanied by any diagram at all. The explanation of this was that the structure of the joke was enhanced by the presence of the diagram, and this structure kept the parts of the joke well articulated, just as any clearly apprehended relational scheme might do with other varieties of material.

Our own results, in the reproductions of the signs, seem to confirm the superiority of recall-value of the better articulated name-sign combination. In either case, it seems that there has to be an active effort on the part of the subject to make the material meaningful. Taking these findings of Harrower's into consideration, there seems little doubt that the grounds of fittingness for certain of the signs used in the present experiment may be described in terms of structural relevance rather than object-association. The words Philosophy, Wrong, Justice, etc., are general, and the basic concepts for which they stand may be supposed to be devoid of all object-detail and particularity. It seems, further, as though the most appropriate sign must not only have a certain relational significance, but also a formal economy and simplicity of the Gestalt kind.

The question of the fittingness of such signs or diagrams to verbal settings appears to raise many problems which go beyond both the experimental and theoretical scope of the present enquiry and into the psychology of thinking. It seems possible, for instance, that this method may offer a means of evaluating the qualitative structure in an individual's verbal intelligence, so that what he has made of his experience can be objectively assessed.

In conclusion, the present results seem to bring out more clearly the nature of the affective fittingness that Bartlett had observed. The choice of a sign as the most fitting in relation to some general verbal setting appears to depend, not so much upon the possibility of associating it with some object, as upon certain formal qualities of simplicity and regularity of outline combined with a schematic, structural relevance to the particular concept. That the "feeling of fittingness" may be primary in the process of choosing, there seems little doubt, and the reason given is often secondary—a rationalization of a more immediate perceptual process.

V

SUMMARY AND CONCLUSIONS

(1) The choice of a sign as fitting a particular verbal setting appears to depend, not only upon the relationship it bears to the other sign or signs with which it is compared, but upon certain formal, structural qualities of a general character.

(2) It seems probable that it is not, therefore, any object-association, such as is apparent in the verbalizations, which is the primary ground for choice, but rather this structural correspondence between figure and verbal setting.

(3) Within the limits of the material used, it seems justifiable to conclude provisionally that the most fitting sign for a particular verbal setting will combine a conventional structure with a general simplicity of design.

(4) It is, perhaps, in bringing out these structural aspects in the perception of signs in relation to words that the present results have made some contribution to the problem that Bartlett's experiment suggested. Where words are involved in the process of Conventionalization, the generalization apparently necessary may involve both simplification and regularization of form.

VI

REFERENCES

- BARTLETT, F. C. (1932). *Remembering*. Cambridge.
 GIBSON, J. J. (1929). The reproduction of visually perceived forms. *J. exp. Psychol.*, 12, 1-39.
 HARROWER, M. R. (1933). Organization in higher mental processes. *Psychol. Forsch.*, 17, 56-120.

(Received 31st January, 1950.)

STIMULUS INTENSITY AND ATTENTION IN RELATION TO LEARNING THEORY

BY

D. E. BERLYNE

(University of St. Andrews)

Three experiments are described which test the hypothesis that the more intense of two stimuli will, *ceteris paribus*, be more likely to receive attention. It is assumed that an objective behavioural manifestation of attention to a given stimulus is a preference for responding to it rather than to another which is present at the same time.

In all three experiments, successions of pairs of visual stimuli interspersed with single stimuli were presented to the subject, and he was instructed to respond to either (by pressing its corresponding morse-key), but not both, in the case of the pairs. The first two experiments reveal significant tendencies to respond to the larger and the brighter stimulus respectively. In the third experiment, there was a tendency, but a statistically insignificant one, to respond to a constant rather than to a flickering stimulus.

It is shown that the attraction of attention by a more intense stimulus follows from Hull's system with the addition of his new variable, "stimulus-intensity dynamism (V)," and it is suggested that it may thus be possible to add attention to the phenomena that can be integrated with an objective behaviour theory.

I

INTRODUCTION

The experiments to be reported in this paper* were part of a series designed to throw some light on the problem of interest. As the writer has pointed out in a previous publication (Berlyne, 1949), one aspect of behaviour to which the term "interest" has been applied is the preference of the organism for responding to a particular stimulus or group of stimuli when confronted with several. Bartlett, for instance, defines "an interest" as "a preferential response towards certain stimuli or objects or topics" (Bartlett, 1939). Moreover, interest has always been regarded as something closely bound up with attention, and one way, possibly the principal way, in which we can judge from overt behaviour which part of the perceptual field a subject is attending to is to note which part is prepotent in determining his responses.

This is, of course, precisely what happens in the familiar introspective experiments on attention. We present, perhaps tachistoscopically, a display which is too complex for the subject to apprehend the whole of it in the short time for which it is exposed. Then we ascertain what parts or aspects of the display he is capable of describing and conclude that those are what he has been attending to. Now, at one time or another he has learnt to name or describe elements similar to all those shown him; each of them has a particular verbal response which it could evoke. But only those to which he has "paid attention" have actually dominated the situation sufficiently to impose on him their own, even if, as the Gestalt school would remind us, the others might not have been without influence. They have, in other words, been prepotent. Similarly, in studies of fluctuating attention the stimulus in question alternately is and is not capable of calling forth its particular response and overcoming competing response-tendencies. But unfortunately, this interpretation in terms of objectively observable phenomena has usually not been used and most experimenters have worked with relatively complex visual material and with

* These experiments were performed at the Cambridge Psychological Laboratory with the help of a Rockefeller Research Grant.

the verbal response which entails so many complicating factors. We have as a result been prevented from linking attention with other aspects of behaviour and seeing the relevance to attention of the principles that recent learning theory has found to underly all forms of response.

To study attention without the uncertainties of the introspective method, it is therefore desirable to devise as simple a situation as possible in which the subject is presented with more than one stimulus at once but is required to respond to one only. Many well-known studies involving the use of two or more simultaneous stimuli have been made, but they appear to fall into one or other of the following categories:

- (i) Which stimulus is prepotent is made to depend on learning and association with reward or punishment (e.g. Yerkes' discrimination-box, Lashley's jumping-stand and Hunter's delayed reaction).
- (ii) It depends mainly on the present motivational state of the subject (e.g. Jenkins and Warner's obstruction-box and Brown's conflict-experiments).
- (iii) It depends on previous verbal instructions (e.g. experiments on distraction).

Now, such factors as learning, motivation and *Aufgabe* undoubtedly influence attention profoundly, but it has been held by many writers and confirmed by common experience that other, more intrinsic qualities of the stimulus, particularly intensity and novelty, also do so (see, e.g. Titchener, 1901). Little has so far been done to study or analyse the effects of these.

The experiments to be reported are concerned with the question of stimulus intensity, as represented by the size, brightness and steadiness of visual stimuli. A series of pairs of stimuli, with single stimuli intermingled with them, is presented to each subject, and, in the case of the pairs, he must respond to one only. The situation is such that the instructions, learning and motivation cannot favour one choice rather than another, and so intensity is free to show its effects.

II

EXPERIMENTS

Apparatus.—The subject sat facing a vertical wooden board, approximately 18 in. wide by 12 in. high, in which there were four square holes at about eye-level, each $1\frac{1}{2}$ in. per side and $1\frac{1}{2}$ in. from its neighbour, in a horizontal line. Just behind each hole was a 3.8 volt torch-bulb, and in front of it a moveable cardboard screen could be interposed. Each screen had a paper-covered aperture and the size, shape and colour of the stimulus could be varied by varying the size and shape of the aperture and the colour of the paper. Intensity could also be reduced by inserting a resistance in the circuit.

Below each hole at the base of the board was a morse-key, whose end projected through the board to show the experimenter, by means of a simple mechanical device, which key was being pressed.

The subject sat just near enough to the board to reach the keys comfortably. Subjects were naval ratings and students.

Experiment 1

Procedure.—The four stimuli used in this experiment were two white circles, approximately $\frac{1}{2}$ in. in diameter, and two larger white circles, approximately $\frac{3}{4}$ in. in diameter. A different permutation was used with each subject to cancel out the effects of position.

The subject was first told that he would be given a few practice-trials. One light at a time was switched on, and he was to respond by pressing down its corresponding key as quickly as possible, releasing it as soon as the light went off.

Then the experiment proper began. Forty-eight stimulus-patterns were given, consisting of six presentations of each of the four stimuli alone and four of each of the

six combinations of two. The order was random, but the same for each subject. The subject was instructed to respond by pressing the appropriate key in the case of a single stimulus but to respond to one and one only when two appeared. "It doesn't matter which, whichever you feel like pressing first!" Only the first and second fingers of the right hand were to be used, and these had to be returned to a certain spot opposite the middle of the board after each trial.

Half the pairs were irrelevant to the purpose of the experiment, since they consisted of two stimuli of equal size. But these and the single stimuli were introduced to increase the number of possibilities and thus make it harder for the subject to prepare himself, making the choice more immediate and spontaneous.

Results.—The only trials whose results are relevant are those with a presentation of two heterogeneous stimuli (i.e. 16 trials per subject).

Mean responses to larger stimulus: 13.27

" " " smaller " : 2.63

Number of subjects: 6.

$$\sigma^2 = 9.33 \quad t = 3.9 \quad v = 5 \quad p = 0.012.$$

Experiment 2

Procedure.—As in Experiment 1, except that the stimuli were two white circles, $\frac{1}{2}$ in. in diameter, and two *dimmer* white circles of the same size, intensity being reduced by the insertion of a 10-ohm resistance in the circuit.

Results.—Mean responses to brighter stimulus: 13.5

" " " dimmer " : 2.5

Number of subjects: 6.

$$\sigma^2 = 8.75 \quad t = 4.15 \quad v = 5 \quad p = 0.009.$$

Experiment 3

Procedure.—As in Experiment 1, except that the stimuli were two white circles, $\frac{1}{2}$ in. in diameter, and two similar circles with *flickering* lights. The flickering was produced by moving the switches.

Results.—Mean responses to constant stimulus: 10.2

" " " flickering " : 5.8

Number of subjects: 10.

$$\sigma^2 = 10.96 \quad t = 1.99 \quad v = 9 \quad p = 0.078 \text{ (not significant).}$$

III

MISCELLANEOUS OBSERVATIONS

The following facts emerged from all experiments:—

- Almost every choice was made, according to introspective reports, automatically, without any effort, thought or deliberation, although the stimuli were generally perceived simultaneously.
- Introspections were frequently belied by the results. Subjects denied tendencies to prefer responding to the more intense stimulus when such preferences were clearly demonstrated by the results, and some asserted tendencies which did not exist. This has serious implications for the use of the introspective method.
- There were no striking effects of position on the responses and, in any case, randomizing prevented any positional preferences from invalidating the results.

The mean numbers of responses to positions were as follows (from left to right from the subject's point of view):

Experiment 1	..	6.3	5.3	7.7	4.7
" 2	..	5.0	5.3	8.0	5.7
" 3	..	4.2	6.6	6.9	6.3
Mean	..	5.0	5.9	7.4	5.7
(22 subjects)					

(All pairs, heterogeneous and homogeneous, are counted.)

(d) Some subjects showed the following tendencies:

- (i) To adopt a systematic procedure, e.g. to respond always to the right-hand stimulus. This was more common with students than with sailors, and the results of such subjects were ignored; as this prevented any effects of the variable to be studied from showing themselves.
- (ii) To respond to the stimulus just responded to or near it.
- (iii) Rarely, to vary responses intentionally.
- (iv) Rarely, to "compromise," by responding to a stimulus (not lit) between the two presented.
- (v) To attempt, almost compulsively, to anticipate where the next stimuli will occur, although it was realized that they were randomized.
- (vi) To feel anxious lest mistakes should be made, although mistakes (i.e. responses to stimuli not lit) were actually hardly ever made.
- (vii) To develop stock responses for each combination. This would presumably have been more in evidence if the experiment had lasted much longer.

IV

DISCUSSION

It was assumed that in these experiments learning and motivation would be kept constant so that the influence of the intensity factor would come out. The principal motivation, whether desire to conform to the experimenter's wishes, anxiety not to make a mistake, confirmatory reaction, etc., was satisfied equally whatever the choice of response. Since the response was made immediately after instructions, the only learning involved was learning to understand language, which can be assumed not to have varied for different stimuli. The results were also independent of any tendency to respond to position and can therefore be taken as due solely to intensity-differences.

Experiments 1 and 2 represent the familiar phenomenon of a larger or brighter stimulus "catching the eye" or "attracting the attention." The results confirm the hypothesis that a more intense stimulus is more likely to be attended to.

Experiment 3 was undertaken to test the common belief that a changing or moving stimulus likewise attracts attention. The results show a tendency, contradicting this expectation, but conforming to the above-mentioned hypothesis, to respond to the steadier stimulus, i.e., that producing more stimulation over a given period. This tendency was, however, not significant, which may be due to a conflict between two antagonistic factors, namely, the effect of stimulus intensity, favouring the steady stimulus, and another factor, favouring the flickering one. The latter is probably connected with the rôle of novelty and change in attention and may depend on a phenomenon which emerged from other experiments and is at present undergoing further investigation.

Now that the hypothesis may be regarded as confirmed, we must consider whether it can be related to fundamental principles underlying phenomena in other fields of psychology. It would appear that it can in fact be deduced from the system that, despite its undoubted inadequacies and provisional character, has so far succeeded in integrating the widest range of psychological phenomena on a rigid hypothetico-deductive basis, namely, Hull's behaviour theory.

Since these experiments were completed, Hull has introduced "stimulus-intensity dynamism (V)" as a new variable in his system (Hull, 1949). It now appears that our results, which were previously inexplicable by his learning theory, represent the effect of V with learning and motivation held constant.

Hitherto, "reaction-potential (sE_R)," an intervening variable on which the latency (s^t_R), probability (p), amplitude (A), and extinction-resistance (n) of responses depend, was treated as a function of habit-strength (sH_R) and drive (D) (Hull, 1943).

But now it is stated to be dependent also on V , which is a variable increasing with stimulus intensity (i). The tendency to respond to the more intense of two stimuli can thus be shown to follow from the postulates of *Principles of Behavior*, with the addition of the new postulates concerning V :

- (i) V is provisionally stated to increase with i (stimulus-intensity) according to the equation

$$V = A (1 - 10^{-b \log i})$$

where A and b are empirical constants.

The more intense stimulus will therefore have a higher V .

- (ii) "Other things constant" (as they are in the conditions of our experiments), $V = {}_sE_R$ (reaction-potential). The response to the more intense stimulus will therefore have a higher ${}_sE_R$.
- (iii) "Momentary effective reaction-potential (${}_s\dot{E}_R$)" depends on ${}_sE_R$, \dot{I}_R (inhibitory potential) and ${}_sO_R$ (an inhibitory factor which oscillates "according to the normal 'law' of chance") (Hull, 1943, Postulate 10). Since the last two factors, \dot{I}_R and ${}_sO_R$, can be taken to be more or less equal for all response-tendencies in our experiments, the one with the higher ${}_sE_R$ will have the higher ${}_s\dot{E}_R$ for most of the time.

The response to the more intense stimulus will therefore have a higher momentary effective reaction-potential (${}_s\dot{E}_R$) more often.

- (iv) "When the reaction-potentials (${}_sE_R$) to two or more incompatible reactions (R) occur in an organism at the same time, only the reaction whose momentary effective reaction-potential (${}_s\dot{E}_R$) is greatest will be evoked" (Postulate 16).

The more intense stimulus will therefore be responded to more often.
—Q.E.D.

The above postulates and their applicability can be checked by further experiments with different intensities. The probability of response to the more intense stimulus can be found for each pair of intensities, and it can be seen whether they are compatible with Hull's formulae. If they stand the test, the way may be open to the formation of something that has been long overdue, namely, an objective theory of attention.

Since the completion of this paper, the postulates concerning ${}_sE_R$ and V have been modified in the direction of greater complexity (in Hull's cyclostyled Memoranda on Behaviour Theory of Nov., 1949, Feb. 7th, 1950, March 6th, 1950). But these changes are not such as to invalidate the above deductions.

V

REFERENCES

1. BARTLETT, F. C. (1939). *Study of Society*. London.
2. BERLYNE, D. E. (1949). "Interest" as a psychological concept. *Brit. J. Psychol.*, **39**, 184-195.
3. HULL, C. L. (1943). *Principles of Behavior*. New York.
4. HULL, C. L. (1949). Stimulus-intensity dynamism (V) and stimulus generalization. *Psychol. Rev.*, **56**, 67-76.
5. TITCHENER, E. B. (1901). *Experimental Psychology*. New York. (Vol. I, Part II, p. 207.)

(Received 1st October, 1949.)

THE IMAGINARY QUESTIONNAIRE

BY

J. W. WHITFIELD

*(From the Medical Research Council Applied Psychology Research Unit,
Cambridge)*

The experiment was designed to throw some light on the statistical problems in the analysis of questionnaire data. Previous work (unpublished) suggested that a simple choice response was partially determined by previous responses; and also that the nature of the determination was changed with changing length of series. A "null" experiment was devised in the form of a questionnaire without any questions, and the distribution of responses was studied with respect to the problems formulated.

The observations are discussed in three sections.

In the statistical discussion an alternative meaning to overall association or dissociation is advanced. This relates association or dissociation to human behaviour in the serial response situation, rather than to qualities of the questionnaire. It is further suggested that association between specific questions should be tested against the association in the whole questionnaire, and an appropriate treatment is indicated.

The observations depart from statistical randomness in certain ways. Answers made up almost entirely of one form of response are given less often than would be expected. Long sequences of the same type of response are relatively infrequent, and sequences of alternation of response are also rare. As the material is "null" it implies that the human concept of randomness differs from the statistical concept.

An attempt is made to define the human concept of randomness. It appears that a series of responses which has a pattern, or for which the subject can postulate a simple "cause" will not be accepted as random by the human subject. This raises problems of a perceptual and cognitive nature. It also has a bearing on the design of questionnaires or experiments involving serial responses.

I

AIM OF EXPERIMENT

THE purpose of this experiment was to throw some light on the statistical problems involved in the treatment of questionnaire data. With purely objective "factual" types of questionnaire many of the problems discussed should not arise, but in those cases where the questions present a judgment problem to the subject, certain effects due to the nature of the presentation may be observed. Questionnaires in the field of public opinion, or those which purport to give evidence in the study of personality or temperament, and laboratory-type discrimination tasks come into this category.

Preliminary observations had led to two hypotheses:

- (1) In a series of questions involving judgment with only two possible responses (e.g. Yes or No), a particular response is partially determined by previous responses.
- (2) Length of series modifies the nature of the effect.

An examination of the statistical concept of independence necessary for the evaluation of the significance of any observations follows naturally from these hypotheses.

II

MATERIAL

As an attempt to produce a psychologically "null" condition, an Imaginary Questionnaire was produced—i.e., a questionnaire with space for questions and for a "Yes" or "No" response, but with no questions. The subjects were instructed to ask themselves questions of the form "Do you prefer to?" and record their answers in the appropriate column for "Yes" or "No." Three groups of subjects were used, the bulk of them being students of an elementary psychology course, a few were more sophisticated in psychological experimentation. Three sizes of questionnaire were used, 6, 10 and 14 questions. It may be argued that this did not provide a null condition in the psychological sense, a point which will be discussed later, but it is as null as anything the experimenter could think of. The really important feature is that within the limits of two responses any further constraints on behaviour are provided by the subjects themselves.

III

OBSERVATIONS AND DISCUSSION OF STATISTICAL PROBLEMS

TABLE I
DISTRIBUTION OF RESPONSES BY QUESTIONS

Question number	Six questions		Ten questions		Fourteen questions	
	Yes	No	Yes	No	Yes	No
1	34	6	35	5	33	7
2	29	11	27	13	30	10
3	25	15	21	19	28	12
4	28	12	29	11	26	14
5	30	10	28	12	24	16
6	29	11	26	14	27	13
7	—	—	23	17	25	15
8	—	—	24	16	22	18
9	—	—	24	16	21	19
10	—	—	20	20	29	11
11	—	—	—	—	27	13
12	—	—	—	—	23	17
13	—	—	—	—	30	10
14	—	—	—	—	29	11

It will be seen in Table I that there are far more "Yeses" than "Noes." This may be due to the normal verbal ordering of the words—Yes, No, and never No, Yes. The effect is most marked with the first "question," where a Yes answer is obtained significantly more often than a No, even taking the general tendency into consideration. The verbal arrangement hypothesis is somewhat strengthened by observations on a small number of subjects who were asked "mentally" to toss up a coin, and record their guesses as to heads or tails. These showed a preponderance of "Head" responses on the first imaginary toss.

Overall Association or Dissociation and the Problem of Independence.

Usually, in the treatment of questionnaire data we wish to discover whether there is any association between the questions, or, specifically, if two or more of

them are associated in any way. There are $\frac{1}{2}n(n-1)$ pairs of questions, where n is the number of questions, and the same number of 2×2 contingency tables can be constructed. Slater (1947) has implied that these are independent contingencies, and has tentatively suggested that the relevant test of overall association in the questionnaire is the summated χ^2 , from all the 2×2 tables with $\{\frac{1}{2}n(n-1)-1\}$ degrees of freedom. In my opinion this is erroneous, as the association between questions 1 and 3 is partially determined by the association, if it exists, between questions 1 and 2, and between questions 2 and 3 (see note on p. 80). We can, however, with safety summate the χ^2 for the $n-1$ independent 2×2 tables concerning the association of one question with each of the others, and test with $(n-1)$ degrees of freedom.

The problem of an overall association, treated as an exercise in factorial analysis by Slater in the same paper can be considered along other lines. Taking the observations from the Imaginary Questionnaires and constructing the matrices of 2×2 tables we find the following (Table II):

TABLE II

<i>Number of questions</i>	<i>Total observed frequency in Yes-Yes (YY) cells</i>	<i>Total expected frequency in YY cells (calculated in the usual manner assuming independence)</i>
6	312	318.475
10	729	740.900
14	1,638	1,621.650

In a 2×2 table relating a pair of questions, if there are more YY than expected it indicates association between the questions, if less it implies dissociation. Table II summates the observations for all pairs of questions.

Thus there is evidence of slight overall dissociation in the 6- and 10-question series. A first general factor of dissociation seems to the writer to be a difficult concept, and an alternative approach is desirable.

If we consider the distribution of responses amongst the subjects we see reasons for these peculiar effects. The total number of YY entries is a function of the distribution of Y amongst individuals. For example, if a subject has said Yes to five questions he will have contributed 10 YY, as five questions make up 10 question pairs. Similarly, if he has said Yes to six questions he will have contributed 15 YY, and so on. Unless the distribution of Y amongst the individuals corresponds to that expected on independence, discrepancies of the nature demonstrated are bound to happen. The distribution on independence assumes that there is a constant probability of giving a Y to a specific question, and that different individuals are merely different trials, with the same probability. The probability of a Y differs between questions, and the observed proportion of Y to a particular question is used as the Y probability for that question. Thus the distribution of Y in trials (i.e. in individuals) is the expansion of p_1, p_2, p_3 , etc., where p_1, p_2, p_3 , etc., are the probabilities of a Y response to questions 1, 2, 3, etc., as derived from the data. The relationship in the questionnaire may be considered, therefore, as partly composed of the true association between questions, and partly of the behaviour of individuals in answering questionnaires.

TABLE III
DISTRIBUTION OF YESES AMONGST INDIVIDUALS

No. of Yeses	Six questions		Ten questions		Fourteen questions	
	Obs. f.	Exp. f.	Obs. f.	Exp. f.	Obs. f.	Exp. f.
0	0	0.01	0	0.00	0	0.00
1	1	0.23	1	0.02	0	0.00
2	0	1.65	0	0.16	0	0.00
3	4	6.16	0	0.86	0	0.02
4	15	12.62	0	2.89	1	0.12
5	18	13.47	5	6.51	0	0.49
6	2	5.86	13	9.90	1	1.54
7	—	—	16	10.10	2	3.63
8	—	—	4	6.61	11	6.46
9	—	—	1	2.50	10	8.67
10	—	—	0	0.40	1	8.70
11	—	—	—	—	9	6.27
12	—	—	—	—	3	3.08
13	—	—	—	—	1	0.92
14	—	—	—	—	1	0.13
Obs. variance ..	0.834		1.597		3.828	
Exp. variance ..	1.158		2.189		3.01	
χ^2	5.670		8.445		12.013	
Degrees of freedom	2		2		3	
P	about 0.05		<0.02		<0.01	

Total $\chi^2 = 25.128$, 7 d.f. $P = <0.001$.

The labour of expansion of P_1 , P_2 , P_3 , etc., can be reduced—

$$P_0 = q_1 \cdot q_2 \cdot q_3 \cdot \dots \cdot q_n.$$

$$P_1 = P_0 \Sigma(p/q)$$

$$P_2 = \frac{1}{2} \{P_1 \Sigma(p/q) - P_0 \Sigma(p/q)^2\}$$

$$P_3 = \frac{1}{6} \{P_2 \Sigma(p/q) - P_1 \Sigma(p/q)^2 + P_0 \Sigma(p/q)^3\}$$

$$P_n = \frac{1}{n} \{P_{n-1} \Sigma(p/q) - P_{n-2} \Sigma(p/q)^2 + P_{n-3} \Sigma(p/q)^3 - \dots + (-1)^{n-1} P_0 \Sigma(p/q)^n\}$$

P_0 , P_1 etc. = probability of an individual giving 0, 1, etc. Yes responses.

Testing the Association Between Specific Questions.

It will be seen from Table III that the expected distribution on independence is not a satisfactory fit. Thus we can say that independence does not hold. We shall consider later why this should be so. The practical problem lies in the 2×2 table, relating two particular questions, and how to calculate appropriate expected frequencies in the cells. To assume independence in the normal way may be misleading, if the question we really wish to pose is "Are these two questions more associated or less associated than other question pairs in the questionnaire?" This question can be considered by distributing the observed number of agreements (i.e. YY and NN) equiprobably amongst the question pairs. Each question pair will therefore be given expected frequencies of YY and NN (and YN and NY) which are functions of the Yes and No frequencies for the questions and also of the overall association or dissociation in the questionnaire.

The total observed agreement between questions (ΣO) is

$$\Sigma \frac{1}{2} a(a-1) + \Sigma \frac{1}{2} (n-a)(n-a-1)$$

where a = number of responses by an individual

n = number of questions.

This becomes

$$\frac{1}{2} mn(n-1) - nY + \Sigma a^2$$

where m = number of individuals

Y = total number of Y responses.

The total agreement expected on independence (ΣE) is

$$1/m \Sigma \{y_1 y_2 + (m-y_1)(m-y_2)\}$$

where y_1, y_2 , etc., are the number of Y responses to questions 1, 2, etc.

This becomes

$$1/m \{Y^2 - \Sigma y^2 + \frac{1}{2} n(n-1)m^2 - m(n-1)Y\}$$

Unless the distributions of Y amongst individuals corresponds to that expected on the independence hypothesis there will be a discrepancy $\Sigma O - \Sigma E$. In order to test the association between a specific pair of questions i and j , this discrepancy will have to be distributed amongst all the question pairs. The agreement expected on complete independence in the specific pair is the expected frequency of YY (i.e., $(y_i y_j)/m$) plus the expected frequency of NN $((m-y_i)(m-y_j)/m)$. The required expectations will differ by an amount " z " from these, " z " being related to $\Sigma O - \Sigma E$.

Before defining " z " it is worthwhile to note in passing that $\Sigma O - \Sigma E$ is related to the quantity S in Kendall's coefficient of rank correlation. It has been shown that S in the 2×2 table is equal to the difference between the cross products, and that the denominator required for τ is $\sqrt{A.a.B.b}$, where A, a, B, b are the marginal totals (Whitfield, 1947). If we were to calculate S for all possible 2×2 tables in the questionnaire we would find that $S = m(\Sigma O - \Sigma E)/2$. The total S' (denominator) is $\frac{1}{2} \{(\Sigma \sqrt{b})^2 - \Sigma b\}$ where b_1, b_2 , etc., are $y_1(m-y_1), y_2(m-y_2)$, etc. Thus, if we wished, we could express the overall association or dissociation as a τ correlation

$$\tau = \frac{m(\Sigma O - \Sigma E)}{(\Sigma \sqrt{b})^2 - \Sigma b}$$

but as this is mixed up inextricably in a lot of partial relationships it does not appear to be capable of a significance test directly.*

Using this information, however, we can calculate the quantity " z ."

$$z_{ij} = \frac{b_i b_j (\Sigma O - \Sigma E)}{(\Sigma \sqrt{b})^2 - \Sigma b}$$

To consider an example. In the 10-question questionnaire questions 5 and 7 appear to be related.

* It is likely that with moderately large numbers both of observers and questions the value $(m-1)(m\Sigma a^2 - Y^2)/(mY - \Sigma y^2)$ will be distributed as χ^2 with $(m-1)$ degrees of freedom. It will be necessary to discover more about the conditions under which this approximation is reliable as a significant test. It is derived from Kendall. "Rank Correlation Methods"—on the problem of " m " rankings, regarding the questions as ranking the observers into two sets of tied ranks.

		Question 7		
		Y	N	
Question 5	Y	12	16	28
	N	11	1	12
		23	17	40

The frequency of NN expected on complete independence is $5.1 \chi^2$ with Yates' correction is 6.313 , which suggests a markedly significant degree of dissociation. But there is evidence in the 10-question questionnaire of overall dissociation, and we therefore wish to see whether questions 5 and 7 are more dissociated than the average.

$\Sigma O - \Sigma E$ is -23.8		
Question	b	\sqrt{b}
1	175	13.23
2	351	18.73
3	399	19.97
4	319	17.86
5	336	18.33
6	364	19.08
7	391	19.77
8	384	19.60
9	384	19.60
10	400	20.00
	<hr/> 3.503	<hr/> 186.17

$$\text{Overall } \tau = \frac{-23.8 \times 40}{186.17^2 - 3.503} = -0.031$$

$$z_{57} \text{ is } -\frac{23.8 \times 18.33 \times 19.77}{186.17^2 - 3.503} = -0.28$$

Thus the new expected NN frequency is $5.10 - 0.28 = 4.82$. This reduces χ^2 to 5.298 . Against this must be set the fact that we have reduced the degree of freedom, presumably to $8/9$. Thus there is still evidence that the two questions are dissociated, whatever that may mean in the imaginary case.

IV

BEHAVIOURAL OBSERVATIONS IN THE IMAGINARY QUESTIONNAIRE

Certain aspects of the subjects' behaviour have already been referred to in relation to the statistical treatment. Along with other features they must further be considered in describing the human "random" performance.

Avoidance of Extremes.

The reader will have noted in Table III that, with one exception, fewer individuals than would be expected gave a large preponderance of one response. (The exception is in the case of the Yes response to the 14-question questionnaire.) This accounts for the reduced variance of the 6- and 10-question series. Some previous experiments

had suggested that the relation of observed to expected variance was affected by the number of questions presented. A short series gave a variance much smaller than the expected, but a long series reversed the effect. This is substantiated by these later observations.

The case of the reduced variance in short series can perhaps be explained by arguing that a marked preponderance of one response is not compatible with the subject's idea of random behaviour—it suggests a "cause," i.e. there must be a reason for departure from an average distribution. If this is true, and there is further evidence which appears to support the idea, then it is not surprising that a 6—0 distribution of responses is avoided, constraining the responses into a more limited range. It is also likely that the amount of imbalance of frequency of responses which will be tolerated by the subject is a proportion of the total available range. Hence the tolerated range will increase linearly with increase in the number of questions. The standard deviation of the expected distribution, however, is increased at the rate of approximately the square root of the number of questions, and therefore the relative absence of individuals with marked unbalanced responses becomes of progressively less effect on the variance of the distribution. There may also be a factor working in the opposite direction. If some individuals are more prone to give Yes responses, and some prone to give No responses, even though both may keep within the "tolerated" range there is, as it were, enough room in the longer series for them to demonstrate their proneness, and hence increase the variance.

The Effect of a Response on Subsequent Responses.

Another approach may be made to the question of the difference between the human idea of "random" and true randomness. The question may be asked: "What effect, if any, has a response on its successor, or later responses?" Lag correlations were calculated (i.e., 1st with 2nd, 2nd with 3rd, etc., then increasing the interval—1st with 3rd, 2nd with 4th, etc., and so on). These are shown in Figures 1, 2 and 3. If there were no series effects—i.e. if the probability of a Yes or No response was completely unaffected by previous responses, we would have correlograms which were straight lines at zero, if there were no differences between subjects, or at a positive value if there were a distribution among subjects of Yes-proneness or No-proneness. If, on the other hand, a Yes response inevitably produced a No response as its immediate successor, and vice versa, the correlogram would oscillate — 1.0, + 1.0, — 1.0, etc.

CORRELOGRAMS

FIGURE 1. SIX QUESTIONS

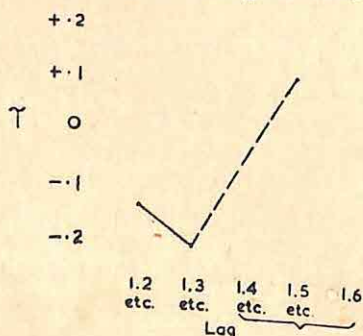


FIGURE 2. TEN QUESTIONS

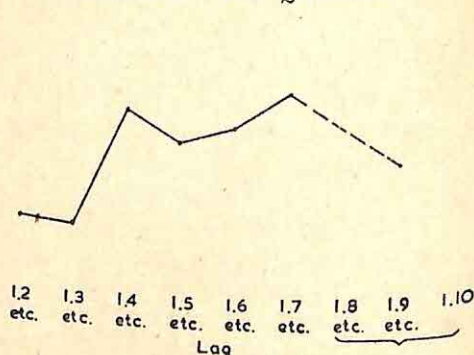
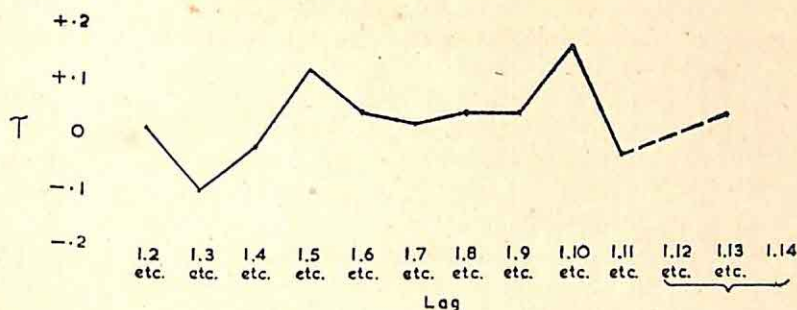


FIGURE 3. FOURTEEN QUESTIONS



It is possible to make some reasonable conjectures as to the behaviour of the subjects from some of the features of the correlogram. Let us consider the 10-question case—the maximum positive correlation occurs with the relations, Questions 1 and 4, Questions 2 and 5, etc. This suggests a series with sequences of not greater than two same responses. A chance series obeying this—i.e. all possible series which do not infringe the limit of two on sequences—gives a correlogram shown in Figure 4. It differs from the observed correlogram in two main respects—(a) more extreme values are obtained, and (b) the difference between the 1.2 type correlation and the 1.3 type correlation is not the same as that observed. A “dilution” of the series—i.e. allowing a few series with sequences longer than two members may account for point (a), and the difference (b) may be accounted for by yet another form of avoidance by the subjects. The lower observed correlation of the 1.3 type implies that subjects have avoided an alternation pattern in their responses. This again fits in with the general idea of avoiding apparent “cause.” A chance series, with 10 questions, allowing no sequences greater than 3, up to

FIGURE 4.

TEN QUESTIONS—CHANCE DISTRIBUTION A
All series within limits of no sequence of same response greater than two.

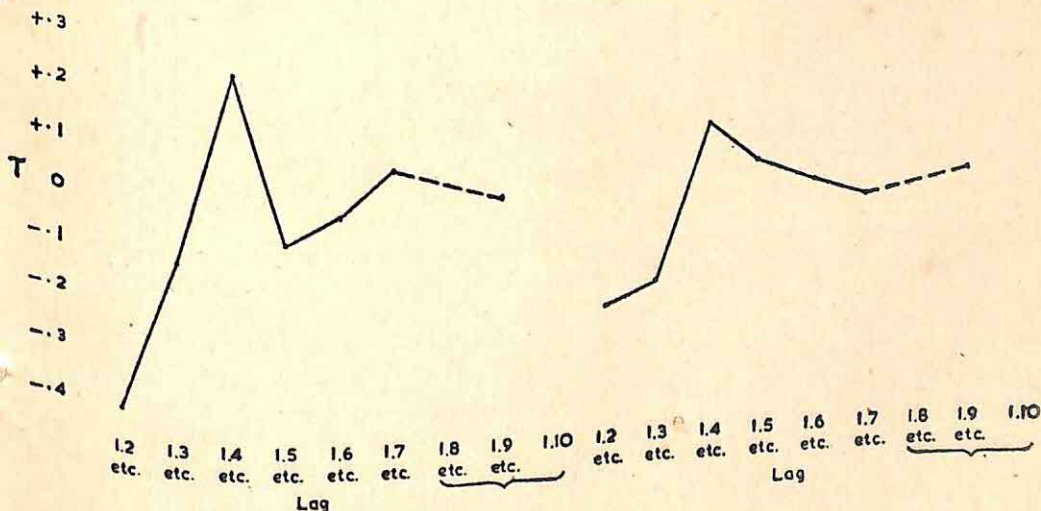


FIGURE 5.

TEN QUESTIONS—CHANCE DISTRIBUTION B
All series within limits of no sequence greater than three, not more than one sequence of three, and no alternations YN YN YN Y or converse or greater.

one sequence of 3, and avoiding alternation sequences of YN₁YN₂YN₃, or converse, or greater, gives the correlogram shown in Figure 5, which in the points considered is moderately satisfactory.

The 14-question series shows a somewhat different pattern, but similar in its main essence—slightly longer sequences would appear to be tolerated, and alternation again avoided. It is also worthy of note that the mid sample of 10 questions from the total of 14 gives a correlogram similar to that of the whole 14, and not to that of the 10 questions, suggesting that the effect is due to the displayed length of series, and cannot be attributed to any marked change of behaviour at the ends of the questionnaire.

The hypothesis concerning the tolerated sequence length may be tested in another way. Given an individual who has shown, let us say, 6 Yes responses to 10 questions—these 6 Yeses may theoretically be distributed anywhere in the 10 questions.

TABLE IV
SEQUENCE LENGTH

	No. of Yeses	No. of cases	No. with sequence of Yeses $> \frac{1}{2}$ Yes frequency	Probability of as few such cases or fewer
14 questions ..	14	1	1	No chance involved
	13	1	1	" " "
	12	3	1	0.2263
	11	9	2	0.01965
	10	1	0	0.6510
	9	10	4	0.6871
	8	11	0	0.09076
	7	2	0	0.5184
	6	1	0	0.8660
	4	1	0	0.8790
10 questions ..	9	1	1	No chance involved
	8	4	1	0.1111
	7	16	7	0.04997
	6	13	3	0.002982
	5	5	1	0.1875
	1	1	1	No chance involved
6 questions ..	6	2	2	No chance involved
	5	18	18	" " "
	4	15	7	0.2231
	3	4	4	No chance involved
	1	1	1	" " "

Taking into consideration only those cases in which there is a chance of obtaining different frequencies, and employing $\chi^2 = -2 \sum \log p$, with 26 degrees of freedom, χ^2 is 48.339, and the combined probability is less than 0.01.

They can be considered as being distributed by the Noes into 5 sequences—first sequence between the beginning of the questionnaire and the first No, second sequence between the 1st and 2nd No, 3rd between 2nd and 3rd No, 4th between 3rd and 4th No and 5th between the 4th No and the end of the questionnaire. The

chance distribution of six objects into five categories is easily calculated. In the lower numerical values of sequences an individual may contribute more than once, and so as a test individuals were divided into two classes—those who gave a sequence of YY longer than half their total YY, and those who did not. The chance of a random series, with the same number of YY showing a sequence greater than half the total Yes frequency was then calculated, and the comparison shown in Table IV. It will be seen that subjects tended to avoid sequences of such length. Unfortunately, a test of the avoidance of alternations would be, at the least, very tedious.

Summing up the evidence we can therefore make the following reasonable conjectures about behaviour in the Imaginary Questionnaire:

- (1) There is a marked tendency to answer Yes to the first question.
- (2) The distribution of Yeses amongst individuals is not chance. It is—
 - (a) Leptokurtic with short series;
 - (b) Less certainly—it is platykurtic with longer series, with a suggestion that the change-over takes place somewhere about the 12-question length.
- (3) Following from (2) even if the purely statistical difficulties of independence in a matrix of 2×2 tables about the questionnaire as a whole are overcome, expected cell frequencies cannot be calculated in the usual manner from the marginal totals, assuming an independence between questions.
- (4) There exist within the series of responses sequential effects which can be described as the avoidance of long sequences of the same response, and also the avoidance of sequences of alternation.

V

DISCUSSION OF THE PSYCHOLOGICAL PROBLEM

Any statement beyond those purely descriptive of findings is, of course, pure conjecture. A few rather wide hypotheses were formulated before these reported experiments were begun. These have been modified, but still remain unverified, and to a large extent difficult to verify. Nevertheless, they may exist as cautions in psychological experimentation. Nature is said to abhor a vacuum—the human subject seems to do so, and objects to thinking about a problem which has insufficient data. Introspective reports from Ebbinghaus-type memory experiments suggest that subjects often attempt to make some sort of “sense” out of the “nonsense” syllables employed. Bartlett’s experiments on social thinking show that even when a considerable amount of information is presented to the subject the normal modes of solution call for the introduction of generalisations or other information not presented in the data and brought in by the subject himself. In the case of the Imaginary Questionnaire the generalization introduced by the subject is that the responses must not imply any consistent “cause.” There is no reason that he should show a relationship between the imaginary questions. The behaviour must then conform to the subject’s idea of relationless, i.e. random behaviour. Several subjects reported that in some cases they had to reverse the form of the question they were asking themselves to avoid repeating the same response more often than they thought desirable. If the observations had shown longer sequences than expected they would have shown consistency of set towards Yes or No in the choice

of questions the subjects produced for themselves—which would perhaps be understandable. The fact that the reverse is true, taken together with the subjects' reports shows that there was an attempt made, consciously by some though probably not all, to randomize the responses. That the responses are not distributed in a statistically random fashion implies a difference between the statistical concept of randomness and the human concept of random behaviour.

The usual human idea of randomness of a series of events is one which shows no apparent cause. This is primarily a perceptual problem. One is unlikely to be dealt four Aces, four Kings, four Queens and the Jack of Spades in a card game, but the same unlikelihood exists for any other specific hand. The former would be perceived and a cause (bad shuffling, good luck, etc.) would be invoked to account for it, whereas in the second case nothing would be thought or said—unless the hand had been specified beforehand. The early writers on probability were concerned with the relation of probability to belief. The question—how often must a penny come down heads in succession from the first before the observer may reasonably doubt its being a perfect penny?—is one which has no generally acceptable statistical solution, even if information is supplied as to the relative frequency of double-headed pennies in circulation. A considerable amount of attention has been given to problems of this nature (see Keynes (1921) for a full discussion). But, except under unlikely condition of being constrained to act in an absolutely logical manner, the usual human response is to make some judgment. After a sequence of events has shown a relationship consistent with a "cause," the observer is inclined to accept the "cause" as demonstrated, or at least to believe he will be more often right than wrong if he so decides. The odd-looking concatenation of events, which would happen if unrelated once in a hundred times is considered as if it were impossible of occurrence by chance, but only by design.

The Imaginary Questionnaire behaviour is in a sense the converse of this; no cause means that there should be no series of events which give the impression of "cause." It would appear that "cause" is attributed to a series which possesses a descriptive simplification, or other distinguishing characteristic. Three such "causes" tended to be avoided by subjects:

- (1) Preponderance of one response over another, which would be taken to imply a mechanism giving preference to one response rather than the other. As no evidence of such a mechanism is otherwise present, this is avoided.
- (2) "Cause" in the sense of long sequences. There is by and large a practical belief in a "law of averages"—or, at a more sophisticated level, a treatment of chance processes as if they were derived from finite and limited series.
- (3) Alternation suggests a simple mechanism of inhibition, or refractory period. Again, failing other evidence in the nature of the experiment to suggest the existence of such a mechanism, the subject avoids alternative responses.

Although it is of interest to know something about the human concept of random behaviour, it is perhaps the relevance of this to the more practical field of experimentation which is important. In experimental design the question of randomization crops up, and the difference between the random tables and the human concept may be important. This is particularly true in those cases where the task is difficult either by absence of sufficient data—as in threshold measurements, or by working near the limits of performance—as in certain forms of intelligence test, or by difficulty of appreciation or apparent irrelevancy—as in many questionnaires used in work on temperament. It would seem desirable to randomize, where possible, in a humanly

acceptable way. An ad hoc solution would be to randomize in short sequences, probably of about 10, giving equal frequencies to correct responses or expected responses. There is also the possibility that the effect could be used directly in certain types of problem. Under what conditions is a subject prepared to behave differently from his concept of randomness in a series which he believes to be truly random? It is possible that this could lead to an objective estimate of certainty of response.

VI

REFERENCES

1. BARTLETT, F. C. (1938). "The co-operation of social groups." *Occup. Psychol.*, **12**, 30-42.
2. KENDALL, M. G. (1948). *Rank Correlation Methods*. London.
3. KEYNES, J. M. (1921). *A Treatise on Probability*. London.
4. SLATER, P. (1947). "The Factor Analysis of a Matrix of 2×2 Tables." *Sup. J.R.S.S.*, **18**, 114-127.
5. WHITFIELD, J. W. (1947). "Rank Correlation between two Variables, one of which is Ranked, the other Dichotomous." *Biometrika*, **34**, 292-296.

(Manuscript received 29th November, 1949.)

ON THE RELATION BETWEEN TRANSFER AND DIFFICULTY OF INITIAL TASK

BY

J. SZAFRAN and A. T. WELFORD*

*(From the Nuffield Research Unit into Problems of Ageing,
the Psychological Laboratory, Cambridge)*

Subjects were required to throw at a target under three conditions of varying difficulty. Six groups of 14 subjects were used, each group receiving one of the six possible orders of presentation of the three conditions.

Results in terms of mean distance from the "bull" indicated that the three conditions were producing different transfer effects. These could not be adequately explained in terms of stimulus- or response-similarity, and accordingly a tentative hypothesis is put forward in terms of the relative difficulty of the condition presented first, to those presented subsequently. It is suggested that transfer tends to be positive from a relatively difficult initial task to a subsequent task which is easier, while transfer will tend to be negative from a relatively easy initial task to one which is more difficult.

The first of the three conditions presented exerted considerably more powerful transfer effects than the second. Three possibilities as to why this should be so, are outlined.

I

INTRODUCTION

TRANSFER of training has usually been treated as a matter of similarity between two tasks as regards either the stimulus presented or the response required or some relation between them, and experiments on transfer have usually taken the form of presenting two tasks in which the second differs from the first in one or other of these respects, and assessing the gain or loss in performance of the second task resulting from performance of the first. "Stimulus" and "response" have in these studies been given connotations of varying width from nonsense syllables and digits to general principles of complex displays and methods of work. But whether they have been detailed or general, the process of transfer has usually been conceived within a framework in fact, if not in name, rigidly associationist, and typified by the use of the term "identical elements" for what is transferred.

It seems clear, however, that this traditional approach deals with a type of carry-over from one task to another which is only a special case of the more general principle that performance at one task affects performance at subsequent tasks in many different ways. It is the purpose of this paper to describe some experimental results indicating another type of carry-over which can be subsumed under the general heading of transfer, but which lies outside the scope of the subject as usually treated.

II

METHOD

The subjects' task was to throw loops of light chain about 3 inches long at a target consisting of intersecting strips of fibre board arranged to form 7×7 3-inch square "boxes," the centre box is being outlined in white to act as the "bull."

* This paper is based on work an account of which was given in N.R.U.P.A. Report No. 3, by J. Szafran, dated June, 1948.

This design of target was chosen not only for the practical reason that the chains fall into it "dead" and do not slide after they have landed, but because it enables the separate scoring of inaccuracies on the "far-near" and "left-right" dimensions. Far-near inaccuracy not only differs from left-right owing to the foreshortening of this dimension as seen by the subject, but would seem to be of different origin. In throwing underhand, for instance, inaccuracy on the left-right dimension is due essentially to a failure to throw in the correct direction, whereas that on the far-near dimension is due either to an error of "elevation" or to releasing the chain at the wrong time, or to throwing too hard or not hard enough, or to some combination of these. The design of target was justified by results as the inaccuracies on the two dimensions behaved very differently in some respects. For our present purposes, however, they were very similar, and in consequence we have for the sake of simplicity added the inaccuracies on the dimensions together when presenting results.

Each subject threw 50 chains under each of three conditions:—

- A. The subject threw directly at the target which was placed flat on the floor some 8 feet away.
- B. The same as A, except that the subject was required to throw over a horizontal bar 32 inches from the ground and 5½ feet away. The target was in full view underneath the bar so that as regards display the task was unchanged. The effector task was, however, made more difficult by the constraint that the subject had to throw in a high "lob" to clear the bar.
- C. The subject threw over a screen which hid the target from direct vision so that it could only be seen *via* a mirror placed beyond it. As the height and distance of the screen were the same as those of the bar, the effector task was the same as that of B. The screen and mirror, however, made the perceptual task more difficult, especially as regards aiming in the far-near dimension which was reversed in the mirror.

The 84 subjects, who came from Polish Resettlement Corps Camps and Hostels in Cambridgeshire, were divided into six equal groups each of which was given one of the six possible orders of presentation of the three tasks. Assignment to these groups was at random, except that they were balanced as regards age.

The subjects were instructed to aim for the 3-inch square which formed the centre of the target and to take their own time. They threw the whole 150 chains at a single session with short breaks between the groups of 50 for the recording of results.

III

RESULTS

The group average results for the three tasks are set out in Table I, from the last two columns of which it will be seen that there are small but consistent differences

TABLE I

MEAN INACCURACY IN INCHES PER THROW FOR THE SIX GROUPS OF SUBJECTS HAVING DIFFERENT ORDERS OF PRESENTATION OF THREE CONDITIONS OF THROWING

Order of Presentation	Condition Presented			Mean of Group	Mean of two Groups having same condition First
	First	Second	Third		
Direct, Bar, Screen ..	5.49	5.08	6.96	5.84	5.93
Direct, Screen, Bar ..	5.68	7.22	5.19	6.03	
Bar, Direct, Screen ..	5.27	4.79	7.31	5.79	5.73
Bar, Screen, Direct ..	5.32	7.10	4.57	5.66	
Screen, Direct, Bar ..	7.13	4.20	4.74	5.35	5.35
Screen, Bar, Direct ..	6.86	4.79	4.41	5.35	
Means ..	5.96	5.53	5.53	—	—

between the various order of presentation groups. The average inaccuracy of the groups who first threw direct is higher than that of the groups who first threw over the bar, and these again are higher than those who first threw over the screen.

This is a surprising result not to be expected from an experiment of this design, which is essentially an expanded form of the well-known transfer paradigm:

<i>Matched Groups</i>	<i>First Task</i>	<i>Second Task</i>
Group 1	A	B
Group 2	B	A

With this design, unequal difficulty of the two tasks will, if it can be represented in the scores as a difference of simple additive factors, make no difference to the totals of the two groups for both tasks together. If, on the other hand, the inequality of difficulty is such as to require representation in the scores by different multiplicative factors, the combined effects of difficulty and improvement with practice will cause the total for the group which receives the more difficult task first to be higher than that for the group which receives the easier first—the opposite of the present finding.

Unless some very remarkable mischance of sampling has occurred, it seems clear that there is an interaction between task and position in the order of presentation. This is brought out by the figures in Table II, from which it will be seen that while inaccuracy in the direct condition was substantially lower when this condition was presented second or third than when it came first, in the screen condition it was a little higher in the second and third places than in the first.

TABLE II
MEAN INACCURACY IN INCHES PER THROW FOR THREE CONDITIONS OF THROWING

<i>Condition</i>			<i>Presented First</i>	<i>Presented Second</i>	<i>Presented Third</i>	<i>All three Positions</i>
Direct	5.59	4.50	4.49	4.86
Bar	5.29	4.94	4.96	5.06
Screen	6.99	7.16	7.13	7.09

Two possible explanations of this interaction appear to deserve consideration:

- that the three conditions are not equally amenable to practice—i.e. practice effects are greater in direct throwing than in the other two conditions,
- that the different tasks produce different transfer effects.

It is impossible to make any adequate assessment of practice effects from the present experiment, but some indication of them for the direct and screen conditions can be obtained from the results of nine additional subjects who threw two separate sets of 50 chains direct, and of 52 additional subjects who similarly threw twice over the screen. The improvements when the second sets were compared with the first were some 7 and 12 per cent. (significant) respectively—i.e. improvement in direct throwing was *less* than in throwing over the screen. This is the opposite of what would be expected if the interaction observed in the present experiment were due to the conditions not being equally amenable to practice. It is true that the practice effect for direct throwing is assessed from the results of only a few subjects and is therefore not very reliable, but even if these are excluded from consideration altogether, the practice effect for the screen condition is greatly at variance with the figures for this condition shown in Table II. It seems, therefore,

that an explanation of our results in terms of different transfer effects produced by the three conditions of throwing is to be preferred to an explanation in terms of differing practice effects. We will proceed to examine these transfer effects in greater detail.

Transfer Between First and Second Tasks.

Comparing the means for the groups of 14 subjects who received each task after each of the others, with the means for the 28 subjects who received the same task first, we obtain figures for absolute and percentage transfer as set out in Table III. The noteworthy points about this table are the substantial positive transfer effects from the bar and screen conditions to the direct condition, and the small *negative* transfer effects from the direct and bar conditions to the screen condition.

TABLE III
TRANSFER EFFECT IN INCHES PER SUBJECT PER CHAIN FROM FIRST TASK
TO SECOND

<i>Presented First</i>	<i>Presented Second</i>	<i>Transfer Effect</i>	
Direct	Bar	+ 0.21	+ 4%
Direct	Screen	- 0.22	- 3%
Bar	Direct	+ 0.79 *	+ 14%
Bar	Screen	- 0.11	- 2%
Screen	Direct	+ 1.39 *	+ 25%
Screen	Bar	+ 0.50	+ 10%

* Indicates significance at the 1% level.

The transfer effect shown in the third column is the difference between the mean for the task concerned of the 14 subjects who received the task in the second position, and the mean for the task concerned of the 28 subjects who received it in the first position.

Importance of the Task Presented First.

TABLE IV
TRANSFER EFFECT IN INCHES PER SUBJECT PER CHAIN FROM FIRST
TASK TO THIRD

<i>Presented First</i>	<i>Presented Third</i>	<i>Transfer Effect</i>	
Direct	Bar	+ 0.11	+ 2%
Direct	Screen	+ 0.04	+ 1%
Bar	Direct	+ 1.01 *	+ 18%
Bar	Screen	- 0.31	- 5%
Screen	Direct	+ 1.17 *	+ 21%
Screen	Bar	+ 0.56 †	+ 10%

* Indicates significance at the 1% level.

† Indicates significance at the 5% level when tasks presented second and third are considered together.

The transfer effect shown in the third column is the difference between the mean for the task concerned of the 14 subjects who received the task in the third position, and the mean for the task concerned of the 28 subjects who received it in the first position.

The indication from the last column of Table I that transfer effects are dependent upon which task is presented as the first of the three is confirmed by Tables IV and V. From Table IV it will be seen that the transfer effects of the task presented first on the task presented third, are essentially similar to those on the task presented second, while from Table V it will be seen that the transfer from the second to the third tasks is in all cases small.

TABLE V
TRANSFER EFFECTS IN INCHES PER SUBJECT PER CHAIN FROM SECOND TASK TO THIRD

<i>Presented Second</i>	<i>Presented Third</i>	<i>Transfer effect when subjects receiving task presented third are compared with all subjects receiving that task second.</i>		<i>Transfer effect when subjects receiving task presented third are compared with those sub- jects who received that task after the same first task.</i>	
Direct	Bar	+ 0.20	+ 4%	+ 0.05	+ 1%
Direct	Screen	- 0.15	- 2%	- 0.20	- 3%
Bar	Direct	+ 0.08	+ 2%	- 0.21	- 5%
Bar	Screen	+ 0.19	+ 3%	+ 0.26	+ 4%
Screen	Direct	+ 0.08	+ 2%	+ 0.22	+ 5%
Screen	Bar	- 0.25	- 5%	+ 0.11	+ 2%

The transfer effect shown in the third column is the difference between the mean for the task concerned of the 14 subjects who received the task in the third position, and the mean for the task concerned of the 28 subjects who received it in the second position. The effect shown in the fifth column is the difference between the mean for the same 14 subjects and 14 subjects who received the task concerned in the second position, both groups having received the same first task.

None of the values are significant.

IV

DISCUSSION

The Cause of the Observed Transfer Effects.

It does not seem possible to account for the transfer effects indicated in Tables III and IV in terms of either stimulus- or response-similarity. On the hypothesis that positive transfer occurs when the stimulus presented by the second task is similar to that of the first we should expect:

- (a) A relatively large positive transfer from direct throwing to throwing over the bar, and vice versa, since the display in these two conditions was approximately the same.
- (b) A relatively small positive or a negative transfer from these conditions to throwing over the screen, and vice versa, since the display in the screen condition was considerably different from that of the others.

With regard to (a), it will be seen from the tables that although the expected positive transfer occurs from the bar to the direct condition, similar transfer in the opposite direction is very small. With regard to (b), the requirements of the hypothesis are fulfilled by the negative transfer from both the direct and bar conditions to the screen condition, but are strikingly contradicted by the large positive transfer from the screen to the direct condition.

On the hypothesis that positive transfer occurs when the *response required* by the second task is similar to that of the first we should expect:

- (a) A relatively large positive transfer from throwing over the bar to throwing over the screen, and vice versa, since the effector action required was the same in both conditions.
- (b) A relatively small or negative transfer from direct throwing to both the other conditions, although not necessarily vice versa, because while the effector action likely to be adopted when throwing direct was inappropriate to the other conditions, that likely to be adopted in these other conditions could be used when throwing direct.

With regard to (a), it will be seen from Tables III and IV that while a positive transfer occurs when the bar condition is preceded by the screen condition, the transfer in the opposite direction is negative—i.e. the reverse of that required. With regard to (b), the expected transfers from direct throwing to the other conditions occur, but while the direction of transfers in the opposite direction are not inconsistent with the hypothesis, their magnitude relative to those discussed in (a) is out of all proportion to what would be predicted on the basis of response-similarity.

The inadequacy of both these hypotheses in the present case, leads us tentatively to suggest an alternative in terms of *relative difficulty* of the tasks. It is that—

- (i) *When an easy task follows a difficult, transfer will tend to be positive.*
- (ii) *When a difficult task follows an easy, transfer will tend to be negative.*

Predictions made on this hypothesis are, of course, subject to modification by the fact that some aspects of the two performances may be similar and lead to transfer along the lines of the conventional transfer theories. This applies especially to positive transfer appearing in the form of "practice effects," which would tend to increase (i) and diminish (ii), so that in a practical case an easy task following a difficult might show a large positive transfer effect while a difficult task following an easy might show a transfer effect which, while still positive, was smaller. The proviso should probably also be made that the difficult task must not be so difficult as to cause failure to achieve any reasonable standard of performance.

On this theory we should expect transfer to be:

- (1) Substantially positive from the most difficult condition, i.e. throwing over the screen, to both the others.
- (2) Substantially positive from throwing over the bar to direct throwing.
- (3) Small or negative from direct throwing to both the other conditions.
- (4) Small or negative from throwing over the bar to throwing over the screen.

It will be seen from Tables III and IV that the pattern of transfer effects observed is adequately accounted for by this theory.

When the scores are expressed in units which permit them to be regarded as the sums of additive factors, it is convenient from certain points of view to state this hypothesis in another way:

If two tasks of unequal difficulty are presented, achievement at the two tasks considered together, after allowance has been made for practice effects, will tend to be greater when the harder is presented first, than when the easier is presented first.

The hypothesis stated in this form agrees with the means in the last column of Table I, where, as has been previously indicated, the mean inaccuracy for all three tasks taken together was least when the subjects threw first in the hardest condition, i.e. over the screen, and greatest when they threw first in the easiest, i.e. direct.

The Nature of the Transfer Observed.

The large and consistent transfer effects from the first task to the second and third, and the small transfer effects from the second to the third, make it clear that the transfer with which we are dealing is not something which follows inevitably and automatically upon performance of any task, but is to be sought in terms of some influence of the *initial* task upon performance at subsequent tasks. As to the nature of this influence, three possibilities appear to deserve consideration:

- (a) *The receptor comprehension and effector organization* necessary for the performance of a task is more easily built up than modified, so that a subject will tend to deal with subsequent tasks by means of the comprehension and organization built up during the performance of the first. A difficult task will produce a greater measure of comprehension and organization than will an easy, so that a subject who has a difficult first task will approach subsequent easy tasks with what is in a very real sense a greater ability than he would otherwise have had. When, on the other hand, an easy task comes first, the comprehension and organization it produces will be inadequate for dealing with subsequent difficult tasks, and will require modification when these are performed.
- (b) The subject approaches the first task with some *expected standard of performance* which is preformed and more or less independent of the condition under which the task is presented, e.g. he regards a bull as "good" and an inner as "fair" and anything else as "poor." When he approaches a second task he recognizes that it is easier or more difficult than the first, and adjusts his standard of performance accordingly. The standard set for an easy task which follows a difficult will thus be higher than that set for the same task when it comes first. Similarly the standard set for a difficult task which follows an easy will be lower than that set for the same task coming first.
- (c) The subject approaches the initial task with some expected standard of performance as described above, and adjusts his *level of care, effort, etc.*, to attain this standard. When the initial task is difficult, this level will be high; when it is easy, the level will be low. If the initial level carries over to subsequent tasks, those which are easier than the initial task will be done better than they would have been had they come first in the series, while those which are more difficult will be done worse.

The evidence from our present results does not permit us to choose any one of these possible explanations to the exclusion of the others. The three types of possibility that we have put forward would seem, indeed, to be complementary rather than alternative, although the mechanisms postulated may be of different importance in different experimental situations. It would appear, therefore, that further work on difficulty in relation to transfer can profitably aim not at establishing the operation of one of these or other similar mechanisms to the exclusion of others, but at assessing quantitatively the extent of the operation of each.

We wish to record that Mr. C. B. Gibbs, working independently of us on a skilled task of a different kind, has noted (A.P.U. Report No. 113, 1949) transfer effects similar to those found by us. Also Mr. W. Adiseshiah and Mr. H. Kay, following Gibbs, have observed similar phenomena with other types of task. We gratefully acknowledge many valuable suggestions from them and from Mr. A. E. D. Schonfield during the preparation of this paper.

(Received 22nd September, 1949.)

BOOK REVIEW

The Biology of Mental Defect. By LIONEL S. PENROSE, M.A., M.D. With a Preface by Professor J. B. S. HALDANE, F.R.S. London: Sidgwick and Jackson, 1949. Pp. xiv + 285. With 7 plates. 21s.

There is Oliver Goldsmith's authority for saying that "were angels to write books they would not write folios"—most reviewers would agree. Professor Penrose, having written only a demy octavo, has not jeopardised his celestial prospects; and his book is exceptional in leaving an appetite for more. It has immediate relevance to at least four subjects—Psychology, Sociology, Medicine, and Human Genetics—and to review it fully would require a panel of specialists. Of his thirteen chapters, six deal with general topics—the historical background; definition, incidence, and classification of mental defect; principles of causation; and the methods of genetic analysis: six with the aetiology and natural history of specific forms of defect; and a final chapter with treatment. To psychologists the book's most obvious value lies in providing a comprehensive but not unwieldy account of mental deficiency, for which both teachers and students can be grateful. Its long-term importance, however, is probably to be found in the point of view, rather than in its practical utility in the class-room.

The word "Biology" in the title is one key to this point of view. Another is the remark that "the study" (of Human Genetics) "forms a descriptive and inductive science with some resemblance to Astronomy as opposed to Experimental Physics." Professor Penrose's attitude, in short, is that of the empirical (almost sceptical) biologist: and this carries with it certain modes of thought which are favourable to the conduct of research. It implies, in the first place, a descriptive, even taxonomic approach to the subject; only when the relevant variables have been defined in this way is exact statistical or experimental treatment possible. It also implies a regard for the whole individual, and for his mental treatment possible. Human beings are the object of study and they are not to be set aside as the incidental repositories of subnormal I.Q.s. Concepts such as Mental Defect, and Intelligence, in their common sense, have a considerable vagueness in definition—which is one reason why they are useful. They are no more measurable than is Health, and to identify them with what can be measured, e.g. performance on a test, is like regarding the clinical thermometer as a means of measuring Health. To borrow Ramsey's words: "the chief danger to our philosophy, apart from laziness and woolliness, is scholasticism, the essence of which is treating what is vague as if it were precise. . . ." Mental tests are seen as useful aids to clinical study, and no more; further, "each different mental test measures a different aspect of intelligence, just as measurements of stature, span, or girth measure different aspects of body size." Failure to perform adequately tasks of this kind is an important symptom, or symptom complex, but there are many other characteristics of defectives to be noted by the biologist. Testing the urine may, for instance, be more useful than testing the intelligence.

In conformity with this naturalistic attitude, is a refusal to be misled by the continuity of some metrical attribute into a belief that associated attributes or their causes must exhibit a similarly continuous variation. The belief that only a distinction of degree can be made between the normal and the mental defective, because both can be fitted into a continuous distribution of I.Q.s. is no better founded in logic than the similar belief that a continuous distribution of scores on a test of "neuroticism," as between normals and neurotics, will allow us to dispense with the distinction between them. Professor Penrose in his consideration of the aetiology of defect is at pains to discover under what circumstances defectives may be regarded as normal variants and when abnormal determinants must be sought. A good example is to be found in the separation of cases of high and low grade defect—admittedly a classificatory rather than a fundamental distinction. Current teaching on this point too often relies on the fact that the distribution of Binet Test scores departs from normality at the bottom end. There are more low grade defectives than would be "expected"—therefore, the argument runs, some of them are aetiologically distinct from the others. The "expectation," however, presupposes that in test scores we have a direct linear measurement of the biological variable. A scalar distortion would equally well explain the observations. In summarising the qualitative differences which are actually found, Professor Penrose puts the distinction on much firmer ground. I have considered points of methodology more explicitly than Professor Penrose treats them in his book and if I misrepresent him, I hope it may serve only to irritate him into

setting out more formally his own point of view. Nevertheless, I think all of this is implicit in his work, and largely taken for granted—as it probably would be taken for granted by most biologists, though perhaps not by most psychologists. *The Biology of Mental Defect* is not a polemical work, and often disapproval is conveyed quietly, but unmistakably by omission. Like Macaulay's conversation, after he returned from India, the work is "enlivened by occasional flashes of silence." A glance through the bibliography will show how many unedifying ideas we have been spared.

The writing is lucid and compact, and often marked by an attractively barbed form of understatement ("Goring expounded the view that . . . the criminal was, on the average . . . inferior to the man who arrested him. . . ." Nineteenth-century criminologists "initiated an enormous amount of measuring of prisoners.") In places the argument is perhaps over condensed, though this seems generally to be the result of a praiseworthy effort to keep the mathematics from proliferating unwholesomely. I doubt, for instance, whether anyone not already familiar with the subject would follow the derivation of the formulæ for consanguineous unions on pp. 100-101. Nevertheless, it is a considerable achievement to have written the chapter on Methods of Analysis in Human Genetics in terms which the non-mathematician should as a rule be able to follow. In the clinical sections he has avoided the gastronomic terminology popular with earlier descriptive writers: even the traditional "cherry-red spot" of the Tay-Sachs macula has been replaced by the more accurate if less succulent epithet "brownish-red." Illustrating these sections are photographs of a really hand-picked assortment of grotesques. The historical background is well described in an introductory chapter; and the opposing attitudes of religious bodies (the contrast lay originally between Eastern and Western Churches) are neatly summarised in Luther's assumption "that idiots were the illegitimate children of the devil," and the later view which christened them "*les enfants du bon Dieu*" (crétin, we are reminded, is a corruption of chrétien). Professor Penrose's own view inclines rather to the latter interpretation, and his chapter on treatment is devoted, encouragingly enough, to the things that can be done to help these children back into the world and not to the methods of assisting them out of it.

Since the book is likely to remain the standard work for a number of years, it is worth noting some points of detail, which might be corrected in a second impression. The distribution on p. 8 is due to Sir James Paget, who published it in the *St. Bartholomew's Hospital Gazette* in 1869; Galton borrowed it (with acknowledgments) 20 years later. The figure given for the proportion of the normal distribution outside 2σ differs slightly, but confusingly on pp. 26 and 27. The distinction between "incest" and "illegal unions" (pp. 57 and 58) is obscure, unless the Professor had in mind (which seems unlikely) the sin of spiritual incest. This, it is true, is not illegal under Common Law; otherwise the terms appear to be co-extensive (bigamous unions are excluded by the context). Affected males, in the notation of p. 164 should be hY not HY . And, finally, the spelling "Rohrschach," though pleasing, is unusual.

A. D. H.

EXPERIMENTAL PSYCHOLOGY GROUP

MEMBERS

BABINGTON SMITH, B., O.B.E., M.A. (Hon. Treasurer)	..	Institute of Experimental Psychology, 34, Banbury Road, Oxford.
CARPENTER, A., M.B., Ch.B. Psychological Laboratory, Downing Place, Cambridge.
CHAMBERS, E. G., M.A. Psychological Laboratory, Downing Place, Cambridge.
CRAWFORD, Miss A., M.A. Department of Psychology, University of Liverpool, Liverpool.
DAVIS, D. RUSSELL, M.D. (Editor) Psychological Laboratory, Downing Place, Cambridge.
DREW, G. C., B.Litt. (President) Department of Philosophy and Psychology, The University, Bristol.
EYSENCK, H. J., M.A. Ph.D. Institute of Psychiatry, The Maudsley Hospital, Denmark Hill, London, S.E.5.
FOSS, B. M., M.A. Institute of Experimental Psychology, 34, Banbury Road, Oxford.
GRINDLEY, G. C., M.A., B.Sc. Psychological Laboratory, Downing Place, Cambridge.
HARRIS, A. D., M.R.C.S., L.R.C.P. Psychological Laboratory, Downing Place, Cambridge.
HALL, K. R. L., M.A., Ph.D. Barrow Hospital, near Bristol.
HARPER, N., Ph.D. National Institute for Research in Dairying, Shinfield, Reading.
HEIM, Miss A. W., M.A., Ph.D. Psychological Laboratory, Downing Place, Cambridge.
HICK, W. E., M.D. Psychological Laboratory, Downing Place, Cambridge.
HIMMELWEIT, Mrs. M. A., Ph.D. Institute of Psychiatry, The Maudsley Hos- pital, Denmark Hill, London, S.E.5.
HOTOPH, N., M.A. London School of Economics, Houghton Street, Aldwych, London, W.C.2.
JOYNSON, R. B., B.Litt. Department of Psychology, University of Nottingham, Notts.
MACKWORTH, N. H., M.B., Ph.D. Psychological Laboratory, Downing Place, Cambridge.
MACNEIL, Miss F. E., Ph.D. Department of Psychology, University of Manchester.
OLDFIELD, R. C., M.A. Institute of Experimental Psychology, 34, Banbury Road, Oxford.
PICKFORD, R. W., D.Litt., Ph.D. Psychology Department, The University, Glasgow, W.2.
SEMEONOFF, B., Ph.D. (Committee) Department of Psychology, The University of Edinburgh.
TRIST, E., O.B.E., M.A. Institute of Human Relations, Tavistock House, 2, Beaumont Street, London, W.1.
VERNON, Miss M. D., M.A. (Committee) Department of Psychology, The University, Reading.
VINCE, Miss M. A., M.A. Psychological Laboratory, Downing Place, Cambridge.
WELFORD, A. T., M.A. Psychological Laboratory, Downing Place, Cambridge.
WHITFIELD, J. W., M.A. (Committee) Psychological Laboratory, Downing Place, Cambridge.
ZANGWILL, O. L., M.A. (Hon. Secretary) Institute of Experimental Psychology, 34, Banbury Road, Oxford.

Visiting Foreign Members

ADISESHIAH, W., M.A. Psychological Laboratory, Downing Place, Cambridge.
RUSSELL, R. W., Ph.D. Bethlem Royal Hospital, Monks' Orchard, Eden Park, near Beckenham, Kent.

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Vol. II

1950

Part 3

PERCEPTUAL ANTICIPATION AND REACTION TIME

BY

E. C. POULTON

(From the Medical Research Council Applied Psychology Research Unit, Cambridge)

To determine the effect of perceptual anticipation upon reaction time, two different types of experiment were carried out. In the first a skilled response had occasionally to be altered at a given point after a variable warning period. In the second the subject had to react to two auditory signals separated by a short time interval which was systematically varied, the second signal being expected or unexpected.

It was found that lack of readiness to respond to a signal, as revealed by a lengthened reaction time, may be due either to the subject not having prepared himself, as he was not expecting the signal; or to the subject not being able to prepare himself in time. Preparation for reacting to the second of the two signals, when both are expected and have to be reacted to, never appears to take more than between 0.2 and 0.4 seconds, as judged by reaction time. On the majority of occasions it appears to be complete in 0.2 seconds. These times are shorter than those usually given, because the extra delay due to incorrect anticipation has been excluded. With intervals of 0.1 seconds or less, delay in the second reaction may be due to the mechanical difficulty of responding quickly enough, especially when the two reactions have to be made in opposite directions.

The finding that lack of readiness might be due to the subject not expecting the signal, and the further finding that preparation took longer when a skilled response had to be extended than when it had to be stopped, both suggest that so-called psychological refractoriness is due to lack of foreperiod in which to prepare for the response, rather than to a "psychological refractory phase" comparable to the refractory phase of nerves. If, by dividing his attention, the subject was able to prepare for his next response while making his previous one, so-called psychological refractoriness could be completely absent.

I

INTRODUCTION

ALMOST as soon as human performance was first studied in the psychological laboratory, it became apparent that the experimental subject's present response was not entirely a function of the present stimulus, but was co-determined by his past experience, previous learning, etc. However, little attention has as yet been given to the influence of the future on present performance. The experimental subject's performance can show two distinct kinds of anticipation. Receptor anticipation, such as is seen in eye-hand span, is a type of anticipation which is dependent basically upon the timing process. In a simple form of perceptual anticipation the subject is ready to respond in a certain way when he receives an expected signal, or actually responds in this way before he receives the signal, because he knows the signal is due, from his experience of the recent pattern of signals. In a more complex form there need be no external signal, the subject merely performing according to some aim, or idea of what is required of him.

The present paper describes a study of the effect of perceptual anticipation upon reaction time. When a number of signals, each of which had to be reacted to, were

presented to the subject successively at irregular intervals of time (e.g. Telford, 1931; Vince, 1948a), it was found that the subject's reactions interacted with each other, the reaction time depending upon the time interval since the last reaction. This finding led Telford to the theory of a "psychological refractory phase," which he believed to be comparable to the refractory phase of isolated nerves. Attempts such as this to explain the complex in terms of the simple, although frequently successful in both physics and physiology, have on the whole proved singularly unsuccessful in psychology. It is becoming clear that the functioning of higher levels of psychological organization require new concepts specific to themselves, which cannot be arrived at merely from a consideration of the functioning of lower levels of organization. The human subject can anticipate, can make specific preparations for the immediate future, both on the sensory and on the motor side, in a way that an isolated nerve cannot.

In Telford's experiment no attempt was apparently made to eliminate the effect on the subject's performance of his expectancy when, after his last reaction, the next signal would arrive. Yet one would predict from Mowrer's work (1940) that the subject, not knowing whether the next signal would arrive after 0.5, 1.0, 2.0 or 4.0 seconds, would usually aim to be ready for the next signal somewhere between the middle two intervals of 1.0 and 2.0 seconds. In which case he would very definitely not be ready when a signal occurred after 0.5 seconds, and his readiness would have fallen off again if the signal was delayed for 4.0 seconds. The curve for predicted readiness would thus correspond closely to the curve for mean reaction times which Telford actually reported. It seems doubtful, therefore, whether his results have any further significance than this.

A similar criticism can be made of Vince's experiment. In addition, Vince presented some pairs of signals very close together, so that before the subject had had a chance to move his pencil upwards, he received a signal which indicated that he had to move it back again. In these circumstances the subject was not told definitely what he was expected to do, whether he was to move his pencil right up and then down again, or to curtail his upward movement as far as he could. (He was simply told to keep his pencil on the line he was following.) Where he did make the complete up and down movement, the very fact of having to make the upward movement before starting on the downward movement, may have delayed unduly the latter when the time interval between the two signals was short. The delays in reacting which Vince found thus appear to be relatively complex in origin. They cannot be accounted for adequately in terms of a single theoretical concept, and in fact she made no attempt to do so.

Although serial reaction times have been studied experimentally in this way, no attempt appears to have been made to study reaction times for the alteration of a complex graded response. Yet it seems possible that certain effects which are very important in everyday life, may be lost by experimental simplification to the level of serial reactions. Thus while one of the experiments described below is of the serial reaction type, but with expectancy controlled, the other uses complex graded responses instead of simple reactions. This is still a simplification of most everyday life situations, but at least it is less so.

II

OCCASIONAL ALTERATION OF A SKILLED RESPONSE AT A GIVEN POINT

This experiment was designed to present the subject with a reaction time type of situation during the performance of a skilled task. For the setting of a skilled task,

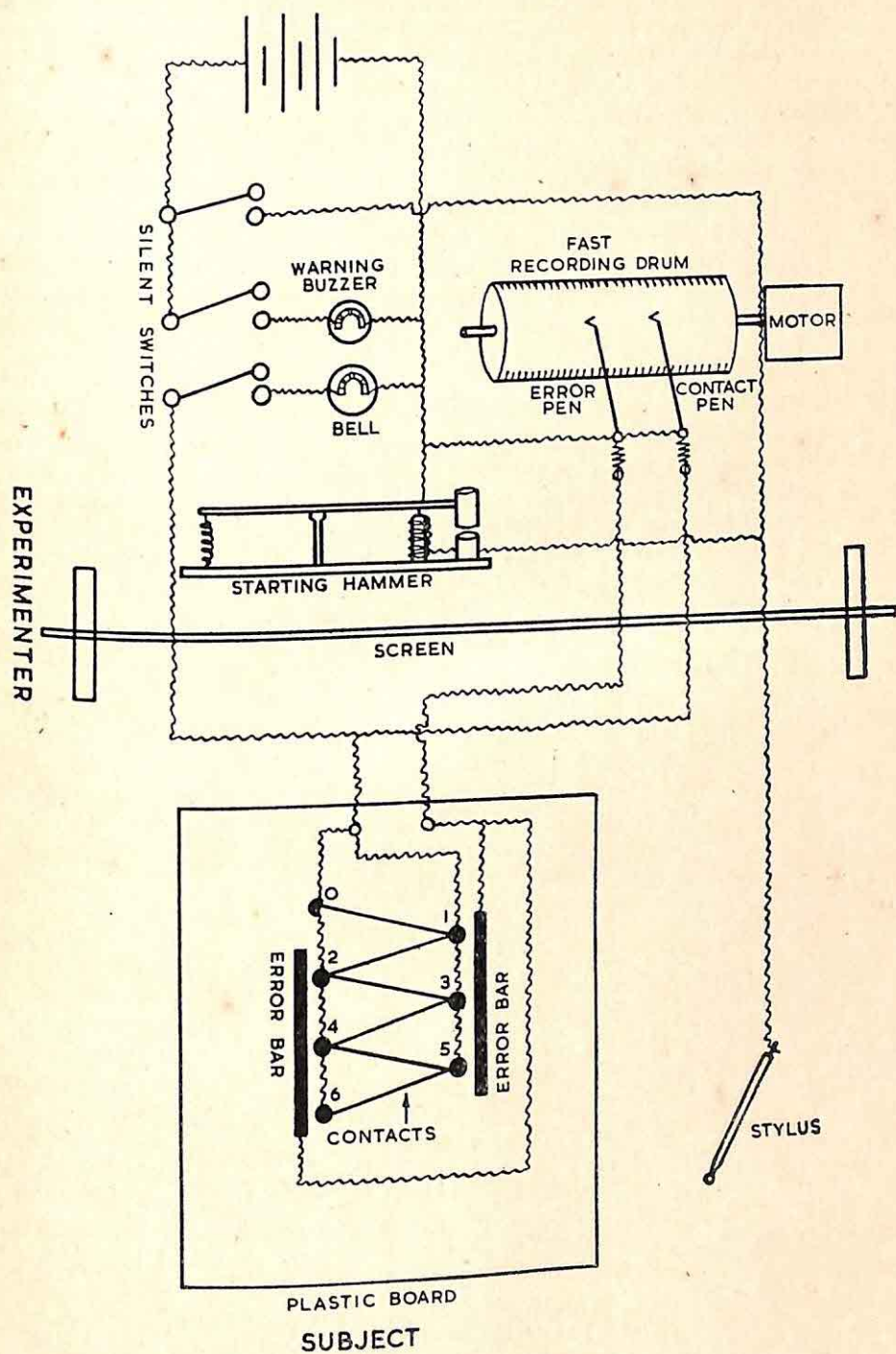


FIGURE 1

The general layout and electric circuits for occasional alteration of a skilled response at a given point.

rather than the classical reaction time experimental arrangement, is closer to situations arising in everyday life. The subject had to trace out a fully visible pattern as quickly as possible. (See Fig. 1.)

The pattern consisted of three identical V's lying on their sides one above the other. Electrical contacts were let flush into a plastic board at the bends, and just beyond these on either side was a brass bar which recorded errors due to overshooting. The subject was given a pencil stylus which he had to keep in contact with the plastic board. At the sound of a hammer, which followed a warning buzzer after an interval of approximately 2 seconds, he was instructed to move off the starting contact point (Contact 0) and trace the pattern as quickly as possible, touching all the other electrical contacts (but not overshooting so as to touch either brass bar), trying to beat a target time he had just set himself. Full knowledge of results as to the time taken, and the errors made in over or undershooting at the bends, was given after each attempt.

On one day a bell sometimes rang during the subject's performance, which indicated that he had to stop on Contact 4 after tracing the first two V's, and omit the third V. On another day the subject was instructed only to trace out the first two V's; in which case, if the bell rang, he was to continue and trace out the third V as well. On a third control day the subject had to pay no attention to the bell when it rang, 12 of the subjects always having simply to trace out the first two V's, the other 12 all three V's. These three experimental procedures, and the lengths of the warning

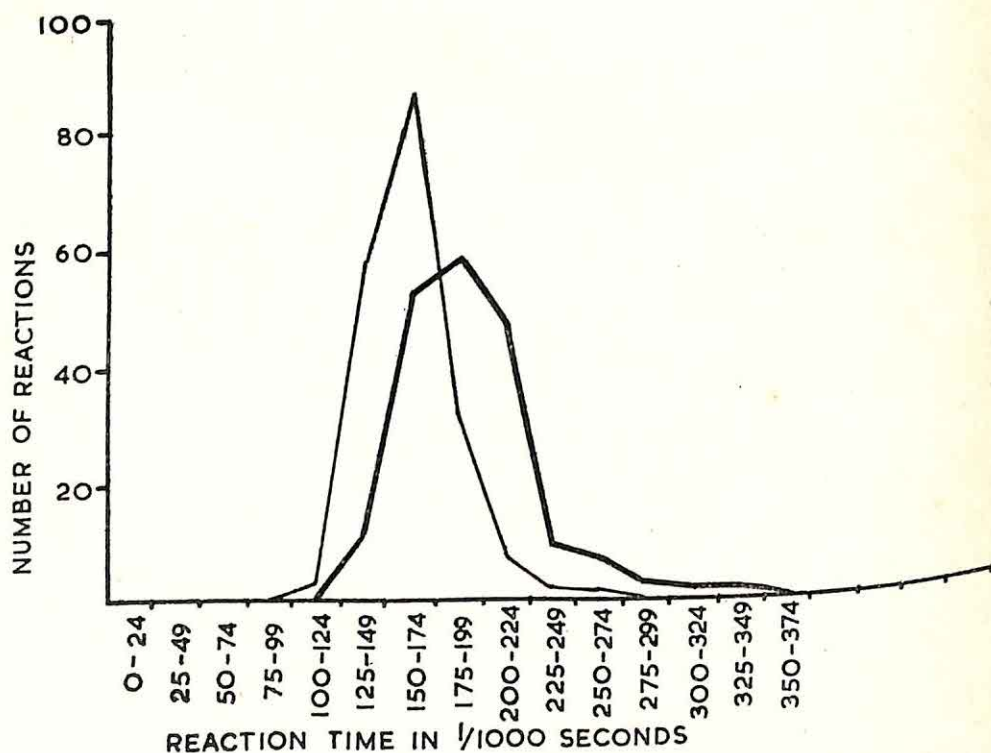


FIGURE 2

The distribution of simple and complex graded reaction times on the third day. Thin line: simple. Thick line: complex graded reactions. Chi squared = 92, compared with 16 at the 0.1% point.

period between the ringing of the bell and the subject reaching Contact 4 (the point at which the performance had to be altered on these occasions), were arranged in Latin Square form. The 24 subjects were all naval ratings, aged between 18 and 26 years.

(a) *Anticipation has a unifying effect in skilled performances.*

It was found that the mean reaction time at the start of the pattern tracing was about 0.025 seconds longer than the subject's simple reaction time when all he was instructed to do was to break the first contact with his stylus, the times lying in the region of 0.175 seconds for the auditory signal given. (See Fig. 2.) This is in line with earlier findings (Woodworth, 1938).

When the subject had suddenly and unexpectedly to change his performance, it was found that he needed a median of roughly about 0.25 seconds longer than an

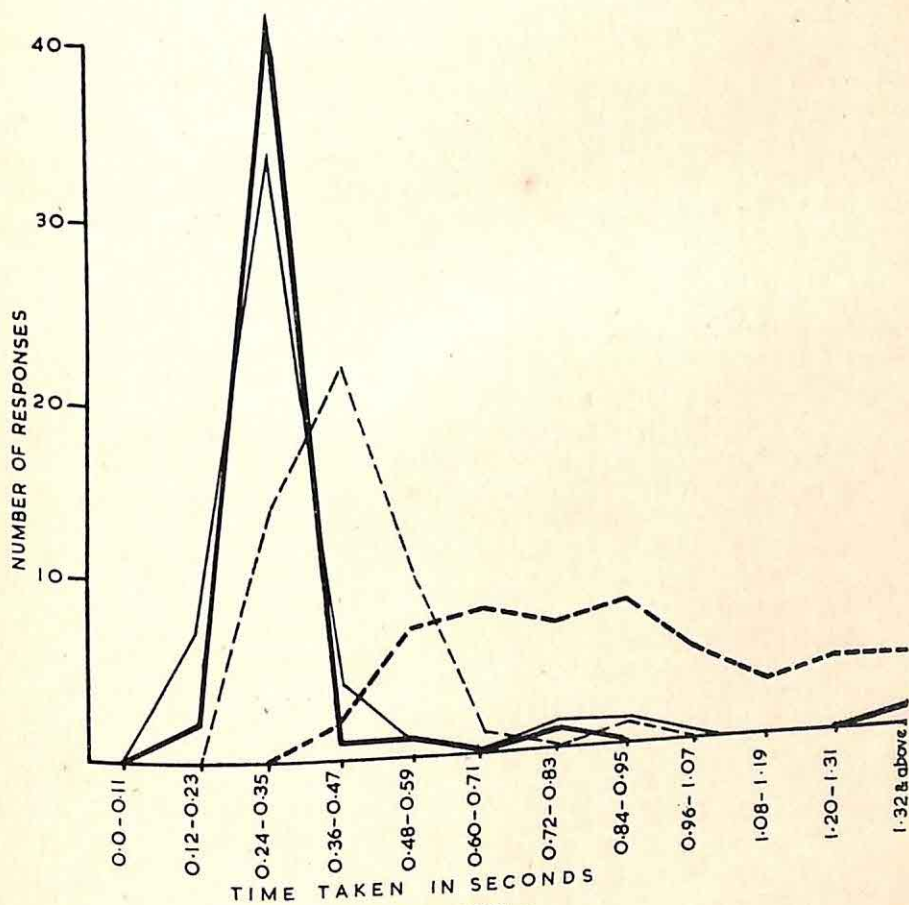


FIGURE 3

The distribution of combined reaction and movement times for the extension of a skilled performance. Thick dotted line: times for a sudden unexpected extension. Thin dotted line: a comparable sample of times from the start of the skilled performance. Thick line: times when a warning for a sudden expected movement of the same size. Thin line: a comparable sample of times when the warning was given before the performance started.

ordinary complex graded reaction time before he could stop the movement, and a median of about 0.35 seconds longer before he could start to extend it, there being significantly more long delays in the latter case. Further, the extension showed significantly more errors, from over or undershooting at the bends, than the rest of the performance, although such an increase in errors was confined to half the subjects. For the sudden unexpected extension, the distribution of times between the bell signal (which occurred as the subject touched what was normally the last contact, Contact 4), and the touching of the first additional contact by the subject, is shown in Fig. 3. Each time thus includes the subject's reaction time and his movement time.

This increase in the initial reaction time of the complex graded or skilled response, as compared with a simple response, and the apparent difficulty in suddenly and unexpectedly altering the skilled response, suggest that even such a simple skilled response as this is not merely a series of discrete responses. The whole performance is dominated by the anticipatory mental set or aim of the subject, and is to this extent a single whole. During the foreperiod the subject is preparing for the whole performance, not merely for the first little bit of it. At the starting signal the subject does not simply move off the first contact point, he begins to make a complex graded movement. And during the performance the subject is making a complex graded movement with a definite point of completion at the end, not merely moving from one contact to the next one. (This does not mean that every detail of the movement has been predetermined during the foreperiod, and that the performance is merely "triggered off" by the starting signal. The results in (b) below, suggest that secondary corrections are possible during the course of the performance, although there will be some delay before they can come into operation).

The performance of the experimental subject is thus not purely a function of the present stimulus and of past experience. It is also partly a function of the future, of the subject's anticipatory mental set or aim. This aim has a unifying effect, converting what might otherwise be a series of discrete responses to a succession of stimulus situations, into a single skilled performance.

(b) *Psychological refractoriness is due to the necessity of creating a new mental set.*

It was also found that provided the bell started to ring at about 0.6 seconds before the point was reached at which the performance had to be altered, reaction time when this point was reached could apparently be eliminated. This is shown in Fig. 3 for a sudden extension of the performance. Presumably the reaction time must have occurred concurrently with the performance of the task, the subject being able to divide his attention between his present performance and the intended alteration. The only occasions on which the performance was slowed down significantly often during the interval between the sounding of the bell and the reaching of the given point, occurred when the task had to be abbreviated. This was only with certain subjects, who generally reported afterwards that they had concentrated more on beating their set targets than on listening for the bell.

When the bell started to ring at about 0.3 seconds before the point was reached at which the performance had to be altered, the results were intermediate between no warning period and a 0.6 second warning period. Thus reaction time was lengthened in about 25 per cent. of the cases, which was only about a third as many as occurred when no warning period was available. Reaction time had apparently been eliminated in about 30 per cent. of cases. There were no significant reaction time differences between extending and abbreviating the movement with this length of warning period.

Practice obtained during the course of the experiment, and the subject's reported expectancy of the bell before starting to trace the pattern, appeared to bear little or no relationship to these findings.

Results such as these can be discussed either in terms of the subject's central nervous mechanism requiring some specific preparation before it can function, i.e. of a necessary foreperiod during which a specific mental set can be built up. Or they can be discussed in terms of some kind of "psychological refractory phase:" for example the recovery of neurosensory or neuromuscular systems, or the necessity of clearing the central nervous mechanism before it can function (Telford, 1931; Craik, 1947). If it is desired to retain any of these interpretations of a "psychological refractory phase," this phase must correspond to the relative refractory phase of nerves rather than to the absolute refractory phase, for the result above shows that it can be eliminated completely when the subject perceives and prepares to respond to a new item while continuing his present response. However, it is possible to take the view that such a division of attention requires a considerable effort on the part of the subject, and is thus comparable to the increase in the intensity of the stimulus which is required to make a nerve conduct during its relative refractory phase.

The conclusion that so-called psychological refractoriness can be eliminated completely for short periods of time when perception overlaps the preceding response, has also been reached from experiments in which the subject was presented with display items successively at high speed. Under these conditions the subject's speed of response was sometimes such that there was no evidence of so-called psychological refractoriness. On such occasions there was sometimes definite evidence that the subject was perceiving one display item while responding to a previous one (Poulton, 1950).

The finding that the reaction time was in general longer when the performance had suddenly and unexpectedly to be extended than when it had to be stopped, suggests that the length of the reaction time was partly a function of the future, of what the subject had to do at the end of it. So-called psychological refractoriness can therefore contain a time element corresponding to the preparation for the succeeding task, such as normally occurs in the foreperiod of the classical reaction time type of experiment. Thus if any of the above interpretations of a "psychological refractory phase" is retained, two theoretical concepts appear to be required, whereas the single concept of a necessary foreperiod is sufficient alone.

But in any case this analogy to the refractory phase of nerves is misleading. For it implicitly assumes that the classical reaction time, a product of an artificial laboratory situation, is the normal state of affairs in the absence of an immediately preceding reaction. But the human subject is never ready to respond as quickly as this, except after a foreperiod; while a healthy nerve is always ready for almost instantaneous response, except just after it has been stimulated. In the classical reaction time type of experiment, the preceding foreperiod leads to the reduction of reaction time to a minimum. It is when no time is available for such a foreperiod that the reaction time is lengthened, and psychological refractoriness is said to occur.

On this view the "psychological refractory phase," a theoretical concept resulting from an approach to human performance which seeks to interpret it purely in terms of past and present, needs to be broadened to take the future into account. The subject delays because he is not ready; but his lack of readiness is due, not to his having to recover from what he has just done, but to his having to prepare himself for what he has to do next, having to create a new mental set. As Bartlett (1947) has suggested, before a task can be performed, mental as well as physical posture has to be appropriate. If this preparation can occur while a previous response is being made, no so-called psychological refractoriness need be apparent.

III

REACTIONS TO TWO SUCCESSIVE SIGNALS

This experiment was designed specifically to discover whether so-called psychological refractoriness was due purely to the subject not being ready for a response because he was not expecting to have to make it, and so had not prepared himself; or whether under certain conditions refractoriness was at least partly due to the subject not being able to prepare himself. For the results of the experiment in §II above (Occasional Alteration of a Skilled Response at a Given Point) suggested that provided the subject had plenty of time in which to prepare himself, and knew when he had to respond (or change his response), he might be able to do it without any so-called psychological refractoriness, regardless of what he had to do immediately beforehand during the foreperiod.

In this simplified version the subject had merely to respond to two auditory signals, a bell and a hammer, the second signal following the first after a short interval of time, by moving a key a short distance in one direction, and then in the reverse direction. The pairs of signals occurred in series of 12, at intervals of 1.7 seconds. In each series the first signal was omitted once, and the second twice, in order to prevent premature reactions. The time interval between the two signals in each pair was reduced from 0.8 seconds in the first series, to 0.1 seconds in the fifth. In each series the subject knew when to expect the signals. On the first day the subject was instructed to respond as quickly as possible to both signals. On a second day, in one part he was instructed to pay particular attention to making his reaction to the second signal as quick as possible; in another part the second signal only occurred twice in each series, and the subject was instructed to try to make his reaction to the first signal as quick as possible. The two parts, two signals, and two response directions were arranged in Latin Square form. After each series, the subject was given knowledge of any false, very early, and very delayed reactions which he had made. The eight subjects were all naval ratings, aged between 19 and 26 years, none of whom had performed in the previous experiment.

The subjects reported that they found the experiment difficult to perform as instructed, one subject actually breaking down and refusing to go on for a while. The difficulty appeared to be associated with the catch episodes. For the absence of the first signal once in each series changed the task from being of simple reaction to being of 2-choice reaction type, the subject having to decide in this case whether the signal was the usual first or second signal, and then to react accordingly. Yet he was instructed to react in general as quickly as possible, which he apparently took as indicating that he was to treat the task as being of simple reaction type.

(a) *Psychological refractoriness can result from disturbances during the foreperiod.*

It was found that when the practised subject knew that the second signal would follow the first at an accustomed interval of 0.2 seconds (except for the customary checks to prevent false reactions), and was instructed to concentrate on keeping his reaction times to the second signal as short as possible, the distribution of these reaction times was significantly different by chi squared from the distribution of his reaction times to the first signal. With all three different sets of instructions, and with all the different time intervals between signals which were used, the distribution of reaction times to the first signal did not differ significantly from the distribution of reaction times to an expected single signal. The distributions for an expected second signal after 0.2 seconds, and for an expected single signal, are shown in Fig. 4.

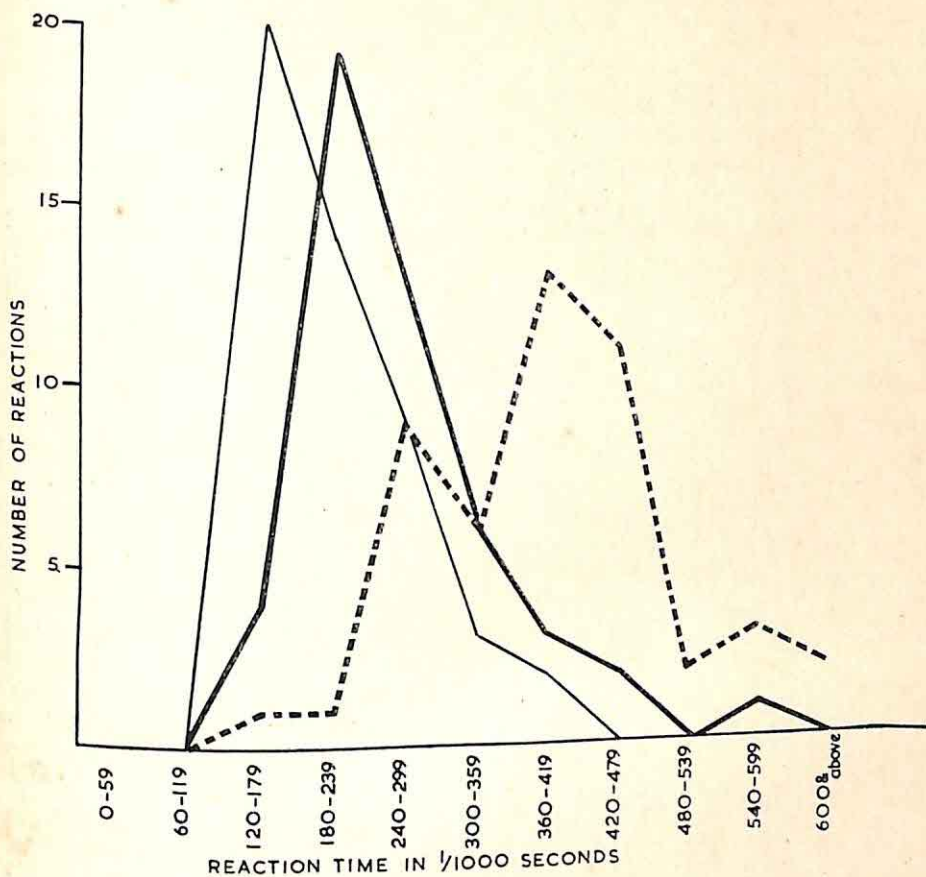


FIGURE 4

Distribution of reaction times to auditory signals on the second day (six reactions from each of eight subjects). Thin line: reaction times to a single signal. Thick line: to a second signal 0.2 seconds after the first, the second signal being expected. Dotted line: to an unexpected second signal.

With this 0.2-second interval between signals there were eight false reactions, all occurring during the 24 catch episodes when one or other signal was omitted. Four of these false reactions occurred on the eight occasions when the first signal was omitted, the subject in each case responding after the second signal as he normally did after the first. He apparently treated the task as being of simple reaction type, responding as soon as he heard a signal, instead of treating it as being of complex 2-choice reaction type, with two different responses according to the nature of the signal. This is confirmed by the distribution of reaction times to the first signal, which does not differ significantly from the distribution of reaction times to a single signal. It was certainly not intended that the catch episodes should alter the nature of the task to being of complex 2-choice reaction type in this way. The surprising finding is that in four out of these eight catch episodes the subjects were able to inhibit their normal response to the first signal. The other four false reactions occurred after the soft click which replaced the usual second signal in 16 of the catch episodes, the reaction times as measured from this click varying from 0.17 to 0.58 seconds. In these cases either the subject was responding at an accustomed interval after the first signal, without necessarily paying any attention at all to the second signal; or he was responding to the soft click, having failed to distinguish between it and the normal second signal. However, in no case in the 96 episodes with this accustomed time interval of 0.2 seconds between signals, did the subject react noticeably too early to the second signal, as might have occurred if he had neglected to

attend to it altogether. Assuming the same incidence in the non-catch episodes, it is probable that in 12 ± 6 of the reactions to an expected second signal, the distribution of the times of which are given in Figure 4, the signal did not receive adequate attention, as judged by the number of false reactions. Whether this was because the subject could not give it adequate attention and also react as quickly as he did, or simply because he did not do so, it is impossible to tell. At all the other time intervals between signals there were less false reactions than at this 0.2-second time interval; and false reactions were rare both when the subject had to react to a single signal, and in the series when the second signal only occurred occasionally.

The results above show that even though the subject knew when to make his second reaction, and could thus in theory prepare himself for it in advance, yet having to react to a signal during this foreperiod a short time before making his second reaction, did delay it on a significant number of occasions. Apparently reactions to two expected signals, either of which on occasions may not occur, requires not a single mental set, but two distinct sets, both of which need time in which to be built up. And part at least of the second set cannot be built up until after the first reaction. A foreperiod of a certain length, undisturbed by having to react to other signals, is thus required if reaction time to a signal is to be minimised consistently.

When the subject was instructed to neglect the first signal, and simply to react as quickly as possible to the second signal, the reaction time distribution was not significantly different from that of an unexpected single signal, both for intervals between the two signals of 0.4 and 0.1 seconds. There were, however, a few more rather long reaction times with the distracting signal, mainly due to one subject. There were also a few more premature and false reactions with the 0.1 second time interval between the two signals, again mainly due to one subject. This was presumably because the 0.1 second interval encouraged the subject to react at the expected time of the second signal, instead of waiting until he had heard it. An expected signal itself during the foreperiod therefore has little if any effect on the reaction to an expected second signal. It is only a second signal to which a reaction has to be made which has any significant disturbing effect.

It will be seen from Fig. 4 that, as judged by the length of the reaction times, the subject was frequently as ready for his second reaction as he was for a single reaction, although subjectively this seemed to require a good deal of effort. This suggests that the kind of so-called psychological refractoriness which cannot be eliminated by forewarning the subject, is only true statistically; that although it holds for a group of responses as a whole, yet it need not occur in every individual case. Readiness as revealed by reaction times, seems to vary from instant to instant. All but the extreme degrees can apparently be achieved relatively quickly, provided the subject knows that it is required of him, and is prepared to make the necessary effort. Simply presenting mean reaction times fails to call attention to this distributional overlap. It also fails to show the skewness of the reaction time distributions, which may invalidate the use of the "t" test in assessing the significance of the differences between them.

When the interval between the two signals was reduced to 0.1 seconds, the delay in the second reaction was much more marked, there being relatively little overlap between the two distributions. But it was found that the distribution of time intervals between the two reactions was not significantly different from the distribution of time intervals which occurred when the subject was instructed to make the double response to a single signal, in spite of there being a few more of the rather longer time intervals between the two discrete reactions. Hence at least part of the delay in the second reaction may have been due to the mechanical difficulty of

responding quickly enough, and not to so-called psychological refractoriness. This is a defect in experimental technique which must almost inevitably occur when the second reaction is the opposite of the first. With such an experimental arrangement measurements of so-called psychological refractoriness must be suspected whenever the two signals are separated by an interval of about 0.1 seconds or less. This defect can at least partly be overcome by making the second reaction an extension of the first (e.g. Vince, 1948b; Hick, 1949).

When the interval between the two signals was increased to 0.4 seconds, the distributions of reaction times to the first and second signal were not significantly different, provided the second signal was expected. Apparently 0.4 seconds gives the subject ample time in which to complete his preparations for his second response, provided he knows he has to do so, and makes the necessary effort. This elimination of the additional delay when an interval of 0.4 seconds was available between the two signals (the second signal being expected), is comparable to the elimination of the additional delay (and apparently of the reaction time also) at the given point ahead in §II (a) when a warning period of 0.6 seconds was available during the performance of the skilled movement. Preparing for a simple choice reaction (to respond or not) at a given time ahead, took about as long as preparing for a definite change of a skilled response at a given point ahead while performing. In both cases the intended response could only receive part of the attention, the rest of the attention being taken up in the first case in determining whether or not the second signal had occurred, for there were occasional checks to prevent false reactions; and in the second case in performing the skilled movement.

(b) *Psychological refractoriness can result from incorrect perceptual anticipation.*

When the second signal only followed the first relatively infrequently, and the subject was instructed to concentrate on keeping his reaction times to the first signal as short as possible, the distribution of his reaction times to the second signal showed relatively little overlap with the first at any of the time intervals between 0.1 and 0.85 seconds. The combined distributions for intervals between signals of 0.1, 0.2 and 0.4 seconds is shown in Fig. 4 above. There was little difference between the reaction time distributions at any of these time intervals.

Fig. 4 shows the considerable lengthening of the reaction time to an unexpected second signal, as compared to an expected second signal when the subject was instructed to react to it as quickly as possible. This and the finding that the reaction time to an unexpected second signal was relatively independent of the time interval between the two signals, suggests that the delay was due to the subject not expecting the second signal when it occurred, and so not having prepared himself for it. Fig. 4 not due simply to the subject not having had time for such preparation. Fig. 4 thus illustrates two distinct factors in so-called psychological refractoriness: lack of readiness due to insufficient time in which to get ready, and lack of readiness due to not realising that there was anything to get ready for. It will be seen that the latter can cause far more delay than the former.

Previous experimenters on the "psychological refractory phase," other than Hick (1948), do not appear to have made this distinction. As usually designed, the subject receives single signals at irregular intervals, with an occasional pair of signals close together. He may not be told what to do if a second signal demands a movement in the opposite direction to the first before he has had time to make the first movement. Sometimes he is not even warned that such pairs of signals are liable to occur from time to time. In such experiments different subjects may be trying to do

different things when a second signal does closely follow the previous one. And it is left to the reader to infer from the experimental design and instructions how far the so-called psychological refractoriness which is reported is due to insufficient time in which to get ready, and how far it is due to the subject not getting ready as he is not expecting a second signal so soon. For example, Vince's conclusion (1948a) that: "A second stimulus is not likely to evoke a second response until an interval of about 0.5 seconds after the appearance of the first stimulus has elapsed," must be restricted to the experimental situation from which it was deduced. As described in paragraph 3 (a) above, when the subject is expecting a second signal, he can always be ready 0.4 seconds after a previous signal to which he has responded. Often he can be ready after 0.2 seconds. Long delays are due predominantly to incorrect anticipation, and not to actual inability to get ready in time.

This finding again demonstrates the inadequacy of discussing so-called psychological refractoriness in terms of a "psychological refractory phase" analogous to the refractory phase of nerves. However long a time the subject is allowed in which to recover from his previous reaction, if he is not expecting to have to make a further reaction, he is unlikely to prepare for one. When a signal further does arrive, therefore, his reaction to it will be delayed.

This additional delay which occurred when the second signal was not anticipated, is comparable to the additional delay which occurred in §II (a) above, when the skilled movement suddenly had to be altered. That the delay was frequently much longer in the case of the sudden alteration of the skilled movement, may have been because here a definite anticipatory mental set had to be discarded, as well as a new mental set adopted; whereas in the case of the unexpected second signal there was no definite mental set to be discarded.

(c) *The effect of simplifying the experimental situation.*

Studying the effect of perceptual anticipation upon reaction time with the simplified experimental arrangement above, which is the only way as yet in which the so-called "psychological refractory phase" appears to have been studied directly, would not have suggested that the psychological refractoriness which occurs even when the subject knows what to expect, is due to his having to prepare himself for his future response, rather than to his having to recover from his previous one. For there could not have been two different responses to make in different parts of the experiment, one requiring a longer preparatory foreperiod than the other, as was found when the skilled response had either to be extended or stopped. This may be a further reason why so-called psychological refractoriness has up until now been treated as a function of the past, rather than as a function of the future; why it has been discussed in terms of a "psychological refractory phase," rather than in terms of a necessary foreperiod.

The findings of these two comparable experimental arrangements thus illustrate Bartlett's (1932) contention that although an experimental simplification may be useful in revealing effects which operate in a complex situation, yet that some effects may be lost by such simplification. One solution of this dilemma would appear to be to experiment as here at the level of both simple and more complex behaviour.

IV

GENERAL DISCUSSION AND CONCLUSIONS

So-called psychological refractoriness contains two distinct elements, which do not seem to have been clearly distinguished by most previous experimenters. Firstly

the subject may not be ready to respond because he is not expecting a signal, and so has not prepared himself. It is this element which seems to cause the long delays before reacting, especially when the subject is expecting to have to complete the skilled task he is performing. And secondly the subject may not be ready because he has not had time in which to prepare himself. Thus if he has had to respond to an expected previous signal 0.2 seconds before the critical signal, his readiness as shown by his reaction time will be less complete on a significant number of occasions than when he has no such previous signal to which to respond. Even in this case, however, he will seem to be just as ready on the majority of occasions. If the subject is allowed 0.4 seconds in which to prepare himself, he can always be ready. These times are shorter than those usually given, because the extra delay due to incorrect anticipation has been excluded. The mere occurrence of an expected signal just before the critical signal has little or no effect when the subject is told to neglect it. With intervals of 0.1 seconds or less between two signals, to both of which a response has to be made, at least part of the delay in responding to the second signal may be due to the mechanical difficulty of responding quickly enough, such as must almost inevitably occur when the second response has to be made in a direction opposite to the first.

Leaving both kinds of so-called psychological refractoriness unseparated, and discussing the combination under the heading of a "psychological refractory phase," is misleading. Firstly, the lack of readiness which is due to a failure on the part of the subject to prepare himself, because he is not expecting to have to respond, is evidently not due to his having to recover from his previous response. And secondly, the time required for such preparation was found to be longer when a skilled response had to be extended, than when it had to be stopped. This suggests that at least part of this time was occupied in the subject preparing for what he had to do next, and not in recovering from what he had just done. It will thus be seen that the analogy to the refractory phase of nerve is misleading. For whereas a healthy nerve is always ready for almost instantaneous response, except just after it has been stimulated, in the intact human subject it is the presence of an immediately preceding foreperiod which reduces reaction time to a minimum. Without such a foreperiod during which the subject deliberately prepares himself, the necessary minimum duration of the foreperiod varying with the task to be performed, reaction time will be lengthened, even in the absence of an immediately preceding reaction. Further, if by dividing his attention the subject is able to prepare for his next response while making his previous one, so-called psychological refractoriness may be completely absent.

The performance of a skilled response is dominated by the anticipatory mental set or aim of the subject. The initial reaction time is longer than a simple reaction time; if the performance has suddenly and unexpectedly to be changed, there is frequently a very long delay before this can occur, much longer than the delay in reacting to a signal which unexpectedly follows the reaction to a previous signal; and if the change in the skilled response involves an extension of the movement, this extension is likely to be erroneous, particularly in the case of certain subjects. The subject's anticipatory mental set or aim thus has a unifying effect, converting what might otherwise be a series of discrete responses to a succession of stimulus situations, into a single skilled performance.

The approach to human performance which attempts to interpret it purely in terms of past and present, thus needs to be broadened to take the subject's mental set towards the future into account. The concept of a "psychological refractory phase" comparable to the refractory phase of nerve, is an example of this backward

orientation on the part of psychologists. Attempts to split up a skilled performance into elements, each of which is believed to depend purely on the elements immediately preceding it, is a further example. The evidence here presented shows that in neither case is this the complete picture. Present performance is always dominated by the subject's idea of the immediate future, by what he expects to happen, and by what he is trying to accomplish.

A comparison of the findings of the two different experimental arrangements used, showed that only the more complex experimental arrangement was able to suggest that the so-called psychological refractoriness which occurs even when the subject knows what to expect, is due to his having to prepare himself for his future response, rather than to his having to recover from his previous one. Although an experimental simplification may be useful in revealing effects which operate in a complex situation, yet some effects may be lost by such simplification.

It is recommended that in any further experiments on so-called psychological refractoriness, the experimental design should ensure that:—

- (i) The subject's expectancy of the signal should be controlled as far as is possible.
- (ii) He should receive definite instructions as to what he is to do when two signals closely follow each other.
- (iii) When intervals of 0.1 seconds or less between the two signals are studied, the required response to the second signal should be in the same direction as the response to the first signal, as in experiments by Vince (1948b) and Hick (1949).

I am greatly indebted to Professor Sir Frederic Bartlett, F.R.S. for presenting the problems and theoretical background out of which these experiments have developed. Also to Dr. N. H. Mackworth, Dr. W. E. Hick, Miss M. A. Vince, and Mr. H. Kay for their help. The experimental subjects were supplied by the Royal Navy. These experiments were carried out while in receipt of a grant from the Medical Research Council.

REFERENCES

1. BARTLETT, F. C. (1932). *Remembering*. Cambridge.
2. BARTLETT, F. C. (1947). The measurement of human skill. *Brit. Med. J.*, June, 1947, 835-838 and 877-880.
3. CRAIK, K. J. W. (1947). Theory of the human operator in control systems. *Brit. J. Psychol.*, **38**, 56-61 and 142-148.
4. HICK, W. E. (1948). The discontinuous functioning of the human operator in pursuit tasks. *Quart. J. exp. Psychol.*, **1**, 36-51. (1949.) Reaction time for the amendment of a response. *Quart. J. exp. Psychol.*, **1**, 175-179.
5. MOWRER, O. H. (1940). Preparatory set (expectancy)—some methods of measurement. *Psychol. Rev. Mono.* **52**.
6. POULTON, E. C. (1950). High speed manual performance and action lag. To be published in Med. Res. Council APU Report series. (Available from the author.)
7. TELFORD, C. W. (1931). Refractory phase of voluntary and associative responses. *J. exp. Psychol.*, **14**, 1-35.
8. VINCE, M. A. (1948a). The intermittency of control movements and the psychological refractory period. *Brit. J. Psychol.*, **38**, 149-157.
9. VINCE, M. A. (1948b). Corrective movements in a pursuit task. *Quart. J. exp. Psychol.*, **1**, 85-103.
10. WOODWORTH, R. S. (1938). *Experimental Psychology*. New York.

(Received 1st May, 1950.)

FIGURAL AFTER-EFFECTS IN STROBOSCOPIC MOTION

BY

L. WEISKRANTZ*

(from the Institute of Experimental Psychology, Oxford)

The present investigation is designed to determine whether figural after-effects of the type reported by Köhler and Wallach occur in stroboscopic motion following prolonged visual fixation of a stationary pattern. The lengths of the paths of two simultaneously presented spots in stroboscopic motion were adjusted so that they appeared equal to each of the 21 subjects. Each subject then fixated a satiation pattern for two minutes, after which he reported whether the paths, unchanged by the experimenter, were equal or not, and in which direction he observed a distortion, if any. The satiation pattern was so designed that a prediction of a distortion in one direction could be made in the subject's first judgment, while a slight change in the arrangement would yield a prediction of a distortion in just the opposite direction in the second judgment. The results are taken to support the finding Köhler and Wallach that the contours of a test pattern recede from the contours of the satiation pattern, even when the test pattern is a stroboscopic one.

I

INTRODUCTION

THE term "figural after-effects" has been used by Köhler and Wallach (1944) to refer to distortions of perceived form which occur subsequent to prolonged stimulation. If, for example, a point on a given pattern (referred to here as the satiation pattern) is visually fixated for a period of about one to two minutes, a pattern (the test pattern) presented immediately afterwards will be found to be distorted in the following manner: (a) contours of the test pattern will tend to recede from the position where the contours of the satiation pattern lay; (b) the amount of this recession, up to a certain limit, will vary directly with the distance from the position of the satiation pattern. The latter relationship is termed "the distance paradox" by the authors. The purpose of the present investigation is to determine whether such prolonged stimulation, or "satiation," will introduce similar distortions into the subsequent perception of stroboscopic motion. The experiment reported here tests only the recession principle (a), and not "the distance paradox."

II

METHOD AND APPARATUS

The method was the comparison of perceived length for two simultaneously presented paths of apparent movement, the comparisons being made before and after satiation. The satiation pattern was so designed that in one case one of the paths should be distorted as so to render it longer than the other, and in another case, just the opposite, i.e. shorter than the other path. The arrangement used was the following:

Conditions were so arranged that T_1 and T_3 were presented simultaneously for a short time followed quickly by the simultaneous presentation of T_2 and T_4 (Fig. 1). Thus stroboscopic motion was obtained so that there appeared to be two spots, one above the other, moving rapidly from left to right. The sequence was repeated a

* Working with a Catherwood Foundation Fellowship.

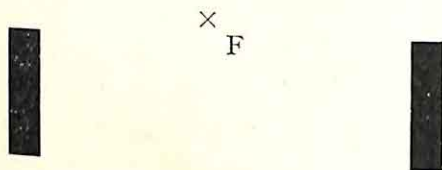
sufficient number of times to meet the requirements of this experiment. The room was dimly illuminated so that the subject could see the fixation point F. While

FIGURE 1



fixating this point, the subject was instructed to report whether the distance through which the top spot moved was larger, smaller, or equal to the distance through which the bottom spot moved. The distances were varied following each report so that they appeared phenomenally *equal* to the subject. After this had been accomplished, the lights were turned on and the following satiation pattern was presented, the subject being instructed to fixate the point F for 2 minutes (Fig. 2). Following this,

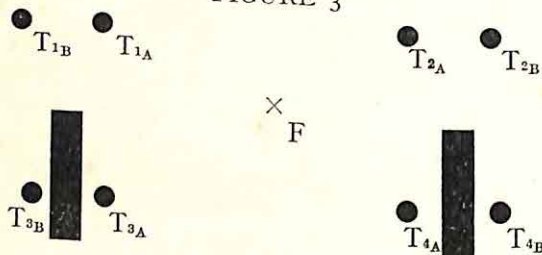
FIGURE 2



the subject once more compared the two distances of the moving spots, although the objective relationships had not been altered by the experimenter.

It will be seen (Fig. 3) that if the spots are arranged so that the distance between

FIGURE 3



them is just shorter than that between the two black bars, and if figural after-effects manifest themselves, the bottom spots should recede slightly from the area of the bars, and hence in the subsequent stroboscopic motion, the bottom path should appear shorter than the top. We shall call this condition A. On the other hand, if the distance between the spots is just larger than that between the bars (condition B), the spots should now recede in the opposite direction, thus making the bottom path longer than the top. Both conditions A and B were presented to all subjects, in some cases with A first, followed by B, and in others with B first. Fixation of the satiation pattern was introduced for both conditions.

The following procedural details should be noted:

1. It was found easier to induce perception of motion if a subject was instructed that he was going to see "two spots moving from left to right across the screen," rather than if the objective stroboscopic relationships were explained to him. In

preliminary experiments, when the subject was aware of this actual arrangement, he often reported that it was very difficult to see motion, but such a difficulty was only rarely present with the above instructions.

2. Fixation of the satiation pattern was monocular, the subject holding a black card over one eye. The test judgment was made with the other eye. In preliminary experiments, it was found that the very vivid negative after-images which appeared after fixation often seriously distracted the subject while making the test judgment. The present procedure reduces the amount of interference, although it also probably diminishes the strength of the effect being sought.

3. It was permissible, of course, only in the first of the two conditions to adjust the distances between the spots so that the paths were phenomenally equal to the subject; such an adjustment in the second condition would counteract the effect being sought. Thus, in the second condition, the spots were arranged so that the distances T_1T_2 and T_3T_4 were objectively equal. The objective distances which satisfied phenomenal equality in the first condition were measured for each subject.

4. While the subject was fixating the satiation pattern, the experimenter made slight irrelevant noises in the general vicinity of the lanterns to make it appear feasible to the subject that the apparatus was being altered. It was discovered in the preliminary investigations that if this was not done, the subject usually expected the two paths to remain phenomenally equal in the test case and, in some cases, even if differences appeared, would refuse to be so obviously "tricked" by the experiment. The subject's attention was further directed to the possibility of change by asking him, while he fixated, "not to mind any noise" behind him.

5. If the subject reported in the test judgment that one path was longer than the other, he was requested to indicate the extent of distortion by selecting the appropriate item in a series of 10 drawings in which the distortion grew progressively greater in steps of 0.5 in., or a step of 2.9 per cent. change in the case of A, and 2 per cent. in the case of B.

The timing mechanism employed to obtain stroboscopic motion was a motor-driven rotating cam closing and opening contacts of two lanterns in the appropriate fashion. The speed of rotation could be varied by means of a rheostat. Generally the time interval between successive presentations was about 0.05 seconds, each lantern remaining lit for approximately 0.18 seconds. The lanterns were approximately 9 ft. away from the screen; the subject sat about 7 ft. away from the screen. In condition A the distance between T_3 and T_4 was close to 7 in. (yielding a visual angle of $4^\circ 46'$); in B this distance was close to 10.5 in. (visual angle, $7^\circ 8'$), the spots falling 0.5 in. outside the contour of the bars in both cases. The distances T_1T_3 and T_2T_4 in all cases were 4.2 in. (visual angle, $2^\circ 52'$). The black bars were 3.3×0.75 in. (visual angles, $\times 2^\circ 14' \times 0^\circ 30'$).

Twenty-one subjects were used in the experiment, most of them being Oxford students. All of the subjects were unaware of the purpose of the investigation, but it was found that many of these students, at least, were fairly suspicious of the nature of the experiment and it was rather difficult to obtain subjects who would adopt a naïve, phenomenonological attitude.

III

RESULTS

The results are recorded in Table I. By a + result is meant one which satisfied the predicted change, while a — result was one which was distorted in a direction

opposite to prediction. Zero indicates that no difference between the two distances was reported.

TABLE I

	+	-	0	P \approx
Combined Results (A and B)	22	3	17	0.000
Average reported distortion ..	5.9%	4.8%		
Condition A	12	1	8	0.002
Average reported distortion ..	7.1%	5.7%		
Condition B	10	2	9	0.019
Average reported distortion ..	4.8%	4.0%		
First reading (12 A, 9 B)	13	1	7	0.001
Average reported distortion ..	6.9%	5.7%		
Second reading (9 A, 12 B)	9	2	10	0.032
Average reported distortion ..	4.8%	4.0%		

It will be noted (Table I) that the differences between positive and negative results are large, and that the probability that such a discrepancy between plus and minus is a chance result is less than 5 per cent. in all cases. The average reported distortion also is greater with the positive results than with the negative, although the difference is probably not significant. A chi-squared test employed to test the hypothesis of chance relationship between conditions A and B yields a chi-square of 11.58, which is well above the requirement for 1 per cent. significance with 1 degree of freedom.

It should be remarked, also, that with no subject was a double negative result obtained, although a double positive result was obtained seven times. In fact, never was there a negative result in which there was not also a positive result obtained for the other condition.

Strictly speaking, it is not crucial that the results be significant for the second reading as well as the first, since there are undoubtedly additional factors of a perseverative nature being introduced when the first judgment is made. It will be noted that the results for the first reading are more favourable than for the second.

In the calculation of probabilities and in the chi-square test, the zero results were ignored. The justifications for such a procedure are as follows:

1. Other figural after-effects experiments have noted considerable individual differences, particularly as regards the size of the distortions. Also, since subjects were asked to describe changes in shading, position, and shape which occurred during the fixation period, it was possible for the experimenter to predict in advance many zero results in those cases where the subjects did not report the characteristic changes during fixation (Köhler and Wallach, *op. cit.*).

2. The resistance which many subjects exhibited towards adopting a naïve phenomenological attitude would most certainly tend to increase the number of zero results. One subject, for example, some days after the experiment confessed that she was not going to be "tricked into saying one path was longer than the other." In addition, it is suspected that the loud clicks emitted by the timing mechanism tended to make subjects suspect that the two distances were objectively equal. More satisfactory apparatus is now being constructed.

3. Some subjects expressed difficulty in making the required comparisons between the paths of motion. It is reasonable to suppose that in such cases the results would most likely be zero.

Finally, it is necessary to establish that the distortions which were reported, could not have been the result of the actual objective discrepancies which were introduced when the subject reported phenomenal equality at the very first stage. Even if these discrepancies were relatively large and in the same direction as the reported distortions, it could still be argued, of course, that the subject nevertheless did judge them to be equal initially and that the distortions are most correctly to be compared with these initial settings rather than with objective dimensions. In order to avoid debate, however, it would be well to consider these discrepancies briefly. Also, it was unavoidable in making the new settings for the second readings to make small errors of adjustment, but in only three instances were these errors greater than 0.1 in. (a 1.4 per cent. error in condition A and a 0.95 per cent. in condition B). These three errors are included in the present considerations.

Considering only those cases in which positive results were obtained, we find that out of 22 cases, in 13 the two distances were objectively equal (to the nearest 1 in.). Of the remaining nine cases, two of the initial discrepancies were in the opposite direction to the reported distortion and hence need not be considered here. The average objective discrepancy for the remaining seven cases was 1.87 per cent., while the average reported distortion was 5.04 per cent. If a *t*-test for paired comparisons is applied, the resulting *t* of 3.96 is significant at the 1 per cent. level, for 6 degrees of freedom. Hence, a hypothesis to the effect that the reported distortions were chance errors in judgment due simply to the objective discrepancies is not tenable. Even if it were tenable, it would affect only seven cases out of the 22, at most (and would also eliminate one of the negative cases). Finally, since the assumptions which such a hypothesis would rest upon are confronted with the observation that the subjects nevertheless did judge these objective distances to be phenomenally equal, it is not felt that it need be taken as presenting a serious challenge to the present interpretation of the results.

IV

DISCUSSION

The present investigation extends the observations of figural after-effects to the case of stroboscopic motion and supports the finding of Köhler and Wallach that the contours of a test pattern recede from the contours of the satiation pattern.

In considering the present experimental results, one might question whether the effect of satiation was, in fact, to alter the path of stroboscopic motion directly, or whether the lower spots T_3 and T_4 were simply altered in the predicted fashion, and the subject was simply comparing the two distances between the two sets of dots. Such a question, of course, rests wholly upon a consideration of whether the subject did or did not, in fact, perceive two dots in simultaneous motion, or whether he merely saw four dots. That in many cases, at least, in which positive results were obtained, the perception of motion was genuine is supported by the fact that many of these subjects were very surprised when informed that there was actually no objective motion taking place. In one case, this disbelief was so strong that the subject was not fully convinced until the experimenter lengthened the time interval considerably and allowed the subject to examine the apparatus carefully. It is fair to assume, on the basis of such observations, that the figural after-effects were directly present in the phenomenal motion and that the experiment is not simply a repetition of a case already fully investigated.

As has already been mentioned, this experiment does not test "the distance paradox," and in that sense is only preliminary. Indeed, it is suggested that any assumption that figural after-effects always present themselves in the same general manner in stroboscopic motion as in the case of other perceptual phenomena already investigated, should be carefully guarded against. Preliminary investigations with other satiation patterns throw some doubt on such an assumption, although it is too early to make a definite assertion. It is suggested, however, that the study of figural after-effects provides a valuable technique for investigating the dynamics of perceptual fields under those conditions in which stroboscopic motion obtains. The technique could easily be extended to study other cases of apparent motion such as the Duncker induced motion type. It would be profitable also, to determine whether continued apparent or real motion of a given type would cause predictable distortions in the subsequent perception of either stationary or moving patterns. In addition, there is some evidence which causes one to suspect that satiation would alter the thresholds for the perception of really and apparently moving patterns. But any attempt theoretically to connect, in the fairly circumscribed area we have selected for this investigation, the dynamics of stroboscopic motion with general perceptual dynamics is obviously premature and awaits more extensive investigation.

The author wishes to thank Professor G. Humphrey for the use of the phenomenon equipment already set up in his experimental room, as well as for his interest and advice. In addition, the author is very grateful to Mr. R. C. Oldfield for his effective aid at many points during the course of the experiment.

REFERENCES

- KÖHLER, W. and WALLACH, H. (1944). Figural after-effects. *Proc. Amer. Phil. Soc.*, 88, 4, 269-357.

(Received 21st April, 1950.)

A NEW PHENOMENON OF APPARENT MOVEMENT AND ITS AFTER-EFFECT

BY

J. A. DEUTSCH

(From the Institute of Experimental Psychology, Oxford)

It is found that a stationary spiral pattern gives an appearance of movement in a flickering light, and, furthermore, that this apparent rotation gives rise to the same kind of after-effect as a spiral actually rotating. The illusion is obtainable over a wide range of conditions. Detailed results are given in the case of six subjects, but a large number of subjects experienced the illusion. The experimenter has as yet found no one who was not subject to the illusion itself. The after-effects are not universally experienced.

I

INTRODUCTION

In 1850, exactly a century ago, Plateau discovered the after-effect of viewing a rotating spiral. He, however, was not the first to note the after-effects of seen movement. In 1834 Addams had published an account of the waterfall illusion. These phenomena attracted considerable attention and gave rise to much controversy, in which Helmholtz and Wundt among others took part. Their discussions are summarized by Boring (1942) and accorded fuller treatment by Wohlgemuth (1911). Controversial interest in this topic has recently revived. With Gibson's (1933, 1937i and 1937ii) discoveries and Köhler's investigations of figural after-effects, attention is again being directed to this field. Fernberger (1948) in a recent note reports an after-effect similar to that of Addams and believes it to be a figural after-effect in three dimensions of the kind described by Köhler and Emery (1947). Also Smith (1948) makes the same identification and criticizes Köhler and Wallach's (1944) satiation theory on this basis.

In the experiment to be reported the dissimilarities between the two classes of phenomena are so marked as to make an identification highly implausible. For it is shown that an after-effect of an impression of movement may occur when the figure of the pattern remains stationary while the ground appears to spin. Köhler's after-effects are on the other hand determined wholly by their figure. The present findings may, it is suggested, also have some bearing on the long-standing problem of the general conditions in which an impression of movement is obtained.

II

METHOD

(1) *Apparatus:* The apparatus used for the experiment was 12-in. white cardboard disk with a black Archimedean spiral on it, unwinding clockwise. This spiral is a curve traced by a point moving uniformly along a line which at the same time revolves uniformly round a fixed point in itself. The lines of the spiral were about one third of an inch thick and three-quarters of an inch apart from each other. For the intermittent illumination the flash of a high power electronic stroboscopic lamp was used. The rate of flashing was continuously adjustable. For the purposes of accurate description, standard conditions were adopted for six subjects. It is felt, however, that some impressions reported by other subjects who saw the phenomenon might be of interest, and they will accordingly be subjoined.

(2) *Standard Conditions*: These were as follows. A dark-room was used with the spiral 7 ft. away from the seated subject. The stroboscope was on the right side of the subject, approximately 6 ft. away from the spiral.

Condition 1: Monocular fixation. The subject was first asked to fixate the centre of the spiral monocularly, i.e. holding a piece of black cardboard before one eye. The period of fixation lasted two minutes.

Condition 2: Contra-lateral after-effect. The subject was thereafter requested to shift the cardboard to the other eye while the stroboscope was switched off and a steady light of approximately equal phenomenal intensity was directed at the disk. The subject was asked to report anything he might see.

Condition 3: Binocular fixation. He was asked to fixate the centre of the spiral binocularly for the same period (two minutes).

Condition 4: Tridimensional after-effect. He was asked to look at a three dimensional object, actually a portrait bust, in steady illumination. The portrait bust was at the same distance as the spiral, but to the right. Thus the subject had to shift his gaze. The rate of flicker which gave optimal movement varied in different subjects from 400 to 800 exposures p.m. (i.e. 6.7 to 13.4 exposures p.s.). Therefore each subject was allowed to fixate the spiral until the optimal rate was discovered. Accordingly the rate of flicker during the fixation period will be reported separately. Movement can in general be seen from 250 exposures p.m. to 1,200 exposures p.m. It has been thought best to give each record separately, on account of interesting variations shown from individual to individual.

In order to avoid the effects of suggestion no leading questions were asked during the course of the experiment. Some of the reports are therefore incomplete.

(3) *Subjects*: These consisted of three students, one male and two female (numbers 1 to 3 in Table I) and three girls, all fourteen years of age (numbers 4 to 6 in Table I). These subjects saw the phenomenon under the conditions described above. Variations in the conditions were made with a larger number of subjects. Any interesting reports from these are included in the general remarks.

III

RESULTS

The main results are given in Table I. The figures give the optimal number of exposures per minute for each subject. Thus, at 600 exposures p.m. the first subject reported white patches rotating with expansion of the spiral and reversing of direction; and slight expansion when the spiral was afterwards viewed with the other eye alone. Binocularly, the impression of rotation was stronger (condition 3). On fixating the bust binocularly he complained of "after image," and then stated that the bust "pulsated (in and out)."

Thus, all six subjects obtained an impression of movement, five reported an after-effect when looking at the spiral itself monocularly after fixation with the other eye in the flickering illumination, and four found an after-effect when they looked at the tridimensional test object (a bust of McDougall). The movement in the after-effect is slow.

General Remarks.—The rotation appears sometimes to include the whole disk and sometimes only the white space. The results seem to show that the impression of movement is enhanced either by binocularly or by length of fixation though both

factors are probably operating. Under the conditions described the length of fixation tires the eye and affects sharpness of focussing. As the pattern used was by no means perfect in its fine details a little blurring might be expected to enhance the illusion. A reduction in light intensity has the same effect.

Many people found the flicker tiring and even unpleasant. The frequencies giving optimal movement were all in the range of the main rhythms of brain activity. The discomfort of the subjects (and the experimenter) may be understood better when we consider that in certain individuals these rates of flashing can give rise to various, forms of severe mental and behavioural derangement as is reported by Grey Walter, Dovey and Shipton (1946) and Grey Walter (1950).

TABLE 1

Subject	Condition 1 (Monocular Fixation)	Condition 2 (Contra-lateral after-effect)	Condition 3 (Binocular Fixation)	Condition 4 (Tri-dimensional after-effect)
1 (male) . . .	600 e.p.m. White patches rotating. Expansion. Reversals.	Slight expansion	600 e.p.m. Rotation stronger. Rest as 1.	After-image. Then slow pulsation.
2 (female) . . .	400 e.p.m. White patches rotating. Contraction. Reversals.	Expansion and approach.	400 e.p.m. Whole disk rotating. Reversals.	Expansion, approach. Then regression.
3 (female) . . .	400 e.p.m. Black rotates, then white.	No after-effect.	400 e.p.m. Polygonal appearance. Complaint of vertigo.	No after-effect.
4 (girl, 14 years)	800 e.p.m. Rotation without reversals.	Expansion.	500 e.p.m. As 1.	No after-effect.
5 (girl, 14 years)	400 e.p.m. Outside rotating, inside still.	Slight expansion	400 e.p.m. Whole disk rotating.	Unsteadiness and movement.
6 (girl, 14 years)	500 e.p.m. Black still. Rotation anti-clockwise.	Expansion. Clockwise movement.	500 e.p.m. Centre going clockwise, outside anti-clockwise.	Bust rocking to and from Cheeks going in and out

Some also reported that the round curves of the spiral became angular. Weitz and Post (1948) and also Weitz and Compton (1950) report a similar transfiguration of a circle into a polygon during prolonged fixation. Those who attempted it could not stop the rotation by taking up an analytic attitude. The movement had for some a quality of unreality. In one case, the only case in which this was done, the movement stopped for the subject when the experimenter gripped the edge of the disk with his hand and started again when he released it. Tilting the disk at an angle did not destroy the illusion. An increase in the rate of flicker produced greater apparent speed of rotation.

IV

DISCUSSION

The after-effects are similar to those of an actually rotating spiral, except that there are reversals in the present case.

For comparison, I quote an account from Bowditch and Hall (1882) of the experiment with an actually rotating spiral.

"In a second series of experiments a large pasteboard disk 22 in. in diameter was inscribed with a black spiral line three-quarters of an in. in diameter, with an equal interval of white and making seven circuits before reaching the circumference. The effects of fixating this disk when in revolution are so striking with some observers as to cause great dizziness and even nausea. . . . If the revolution is in the direction in which the spiral line approaches the centre of the disk the entire surface of the latter seems to

expand during revolution and to contract after it has ceased; and vice versa if the movement of revolution is in the opposite direction. If in the former case the eyes of the observers are turned from the rotating disk towards any familiar object . . . the latter seems to contract or recede in a somewhat striking manner and to expand or approach after the opposite motion of the spiral" (1, p. 299).

We must now consider explanations of this illusion.

1. An eye-movements explanation is contradicted by two facts. First, the movement persists when the flicker rate is too high to permit eye movements during dark periods. Secondly, the illusion does not decrease with proximity as much as might be expected on this explanation. On the contrary, proximity makes very little difference. This is surprising on account of the improved visibility of micro-structure, which furnishes a contradictory cue.

2. It may be argued that the different refractory periods of retinal elements cause these to be stimulated in turn by a rapid flicker, thus setting up a rotating process. But it seems implausible that the retinal conditions should be such as in every case to give an impression of smooth continuous movement.

3. The writer would favour a more complicated analysis of the phenomenon. This rests on two facts. First, when the eye is at rest a movement is perceived when there is a quick simultaneous on-and-off stimulation of adjacent groups of retinal cells. But when the eye is moving no movement is perceived, though there is a quick simultaneous on-and-off stimulation of retinal elements. It is therefore inferred that there exists a central mechanism which co-ordinates information from the vestibular apparatus and the extrinsic eye-muscles and from the eye itself. When there is a quick on-and-off stimulation of retinal elements and a compensating stimulation of the vestibular apparatus and the extrinsic eye muscle receptors, the two sets of signals are made to cancel out by the central mechanism. Now real movement and apparent movement have one important item in common. This is a quick simultaneous on-and-off stimulation of adjacent groups of retinal cells not compensated by another definite stimulation of the vestibular apparatus and the extrinsic eye-muscles. Thus no cancellation takes place and we have a direct impression of movement. Under normal conditions it does not operate unless there is an image actually traversing the retina, for obvious reasons. In the present case, however, a quick simultaneous on-and-off stimulation is produced by the flicker; similarly in the phi- and related phenomena.

Second, what is felt to move, and how it is felt to move, is determined by other features of the situation, namely the spatial framework and the properties of the objects seen. Movement will be seen on a body with successive parts similar in shape and size, so that it will not be too difficult to substitute one for another. In the case of the spiral we would thus expect an experience of rotation, this being the most "plausible" movement for such an object.

To test this suggestion, further observations were made. A plain white solid roll or disk of ticker tape was suspended perpendicularly, first of all with only the edge visible, under conditions similar to those in which the spiral was shown. The disk or roll was presented edge on, i.e. with the narrow convex surface towards the observer. The disk appeared to rotate in three dimensions reversing fairly frequently for some observers. Secondly, the disk was perpendicularly suspended with both the edge and the body of the disk visible. The disk was also seen to rotate in this position. If a piece of ticker tape is left hanging loose, and visible to the subject in this latter position, the outermost sheath of the roll appears to be stationary, while

the main body of the roll rotates. It is also possible to produce an illusion of rotation using a pattern consisting of concentric circles. It seems that these findings support the explanation of the apparent movement which has been outlined above.

I wish to express my thanks to Professor G. Humphrey for the facilities afforded me and to Mr. O. L. Zangwill and Mr. R. C. Oldfield for their kindness and encouragement.

REFERENCES

1. BORING, E. G. (1942). *Sensation and Perception in the History of Experimental Psychology*, 591-593. New York.
2. BOWDITCH, H. B. and HALL, G. STANLEY (1880-82). Optical illusions of motion. *J. Physiol.*, **3**, 297-307.
3. FERNBERGER, S. W. (1948). Figural after-effect in the third dimension of visual space. *Amer. J. Psychol.*, **61**, 291-293.
4. GIBSON, J. J. (1933). Adaptation, after-effect and contrast in the perception of curved lines. *J. exp. Psychol.*, **16**, 1-31.
5. GIBSON, J. J. and RADNER, M. (1937i). Adaptation, after-effect and contrast in the perception of tilted lines. *J. exp. Psychol.*, **20**, 453-67.
6. GIBSON, J. J. and RADNER, M. (1937ii). Adaptation, after-effect and contrast in the perception of tilted lines. II. Simultaneous contrast and the areal restriction of the after-effect. *J. exp. Psychol.*, **20**, 553-69.
7. KÖHLER, W. and EMERY, D. (1947). Figural after-effects in the third dimension. *Amer. J. Psychol.*, **60**, 159-201.
8. KÖHLER, W. and WALLACH, H. (1944). Figural after-effects. *Proc. Amer. Phil. Soc.*, **88**, 269-357.
9. SMITH, K. R. (1948). The satiation theory of figural after-effects. *Amer. J. Psychol.*, **61**, 282-285.
10. WALTER, W. GREY (1950) in *Perspectives in Neuropsychiatry*, ed. Richter, D. London.
11. WALTER, W. GREY, DOVEY, V. J. and SHIPTON, H. W. (1946). Analysis of the electrical response of the human cortex to photic stimulation. *Nature*, **158**, 540-541.
12. WEITZ, J. and COMPTON, B. (1950). Further stereoscopic study of figural after-effects. *Amer. J. Psychol.*, **63**, 78-83.
13. WEITZ, J. and POST, D. (1948). A stereoscopic study of figural after-effects. *Amer. J. Psychol.*, **61**, 59-65.
14. WOHLGEMUTH, A. (1911). On the after-effect of seen movement. *Brit. J. Psychol. Monogr. Suppl.* 1.

(Received 13th May, 1950.)

THE MULTIPLE CRITERION TECHNIQUE OF SUBJECTIVE APPRAISAL

BY

R. G. HOPKINSON

(*From the Building Research Station, D.S.I.R.*)

A technique is described by which the quantitative connection between physical aspects of a stimulus situation and subjective aspects of the perceptual response may be conveniently and reliably determined. The subject is given control of one of the significant physical variables such, for instance, as brightness, and is asked to set this variable to correspond in turn with a limited number of defined criteria relating to a subjective variable such as glare-discomfort. It is found that each criterion acts as a check upon judgments made in terms of the others, so that the scatter of the control settings is less than when a single criterion is used. The functional relationship between the physical and the subjective variable can be estimated, and provided care is taken in the design of the experiment and in the selection of observers, consistent results are obtained.

This technique has been applied during the past ten years to a wide range of visual problems which include those of the visibility of radar echoes, the visibility of street-lighting from the air, discomfort-glare and ease of reading. It is thought that it might find wide application not only in applied psychological work but in the investigation of problems of theoretical import.

I

INTRODUCTION

THE method to be described arose from a study of discomfort glare in lighted streets initiated in 1937. At that time the available methods for the appraisal of glare in this connection were wholly unsatisfactory. It could be demonstrated by direct experiment, and inferred from the Holladay-Stiles formula (Holladay 1926 and Stiles 1929) that *disability* effect of glare from current designs of street-lighting installations was negligible. But it was clear that the discomfort effects could not be ignored. Holladay (1926) had limited his treatment of discomfort to the "shock" effect of a momentary glimpse of a source, and his characterization of glare-conditions was meaningless in relation to street-lighting. Some attempts to appraise glare-discomfort were being made by lighting installation designers, but these were found to be insufficiently consistent for quantitative treatment. The use of arbitrary scales of discomfort did not in itself improve consistency of judgment.

Experience with a laboratory model experiment in which the various relevant physical variables were under independent control suggested that the vital conditions of obtaining consistent judgment were (1) that the observer should *not* be made to assess the degree of discomfort produced by a given stimulus-situation directly in terms of categories or numerical rating-scale; (2) that, on the contrary, he should be asked to make settings of a physical variable to correspond to a limited number of related criteria of discomfort. It was found that this was a task which the observers could readily and willingly carry out, by contrast with their hesitancy and unwillingness over the inverse method. Moreover the use of a number of related criteria was found to refine judgments in terms of each. Considerable care in the choice of criteria was necessary, however, and it was found that the physical variables were not susceptible of equal ease of adjustment. But the method afforded a quantity

of reliable information as to the factors involved in glare-discomfort, and this information was instrumental in permitting the successful re-design of a number of lighting installations.

Later implications of these principles have borne out its usefulness. Such applications have been made to studies of the visibility of radar echoes (Hopkinson, 1946), the visibility of war-time street-lighting from the air (Waldram and Stevens, 1946), discomfort-glare arising in the lighting of buildings (Petherbridge and Hopkinson, 1950) and the relationship between visual acuity and ease of reading under school conditions (Hopkinson, 1949).

II

METHOD

The principles of the multiple-criterion technique, and the precautions necessary for its use, were established by trial and error during the course of the glare investigation. The present procedure is summarized below.

Pilot Experiment: It was found that a pilot experiment was always needed to find out which of the variables could best be put under the control of the subject. For example, the relationship between glare-source intensity and surround brightness for constant degree of discomfort glare could theoretically be studied by causing the subject to vary either the source intensity or the surround brightness. In practice, observers felt much more confident when the source intensity was held constant and their task was to adjust the surround brightness for given criteria of discomfort, than when they were given control of the source intensity. Their settings also showed somewhat smaller scatter in the former than in the latter case.

The reasons why observers preferred to control one variable rather than another, and showed more confidence and less scatter in doing so, were not clearly apparent. An investigation of these factors would be of great value.

Control Experiment: It was also found desirable to have a control experiment extending over the whole period of the investigation. In this the subjects made their assessments at regular intervals (twice daily, for example) on a set of standard conditions. This control experiment had a number of functions. In the first place, it enabled the subject to gain confidence in his ability to make the assessments. Secondly, it provided the experimenter with the opportunity of detecting "off-days" when a subject's settings would wander significantly away from their mean position. Thirdly, it permitted an analysis to be made of the scatter of any subject's observations under standard conditions, so that confidence limits could be derived from his readings under similar conditions. Figure 1 shows a typical plot of one observer's observations under standard conditions in assessment of discomfort glare (Petherbridge and Hopkinson, 1950). The standard deviation of this set of observations is 0.16 on the log. foot-lambert scale, and the coefficient of variation is 0.45.

In general, it will be seen, this observer keeps to his mean assessment fairly consistently, but occasionally wanders well away from the mean (for example on the 18th, 36th and 73rd days). On such days it was considered inadvisable for him to make assessments in the experimental series, especially if his observations were to be pooled with those of other observers.

Selection of Observers: At an early stage of the glare experiment it was decided to conduct the more difficult parts of the investigation with the aid of those observers only whose settings were reasonably consistent. Since, in this experiment, a brightness ratio of 4:1 resulted on the average in a change of one step in the criterion of discomfort, it was decided that an observer should be rejected if he made a setting

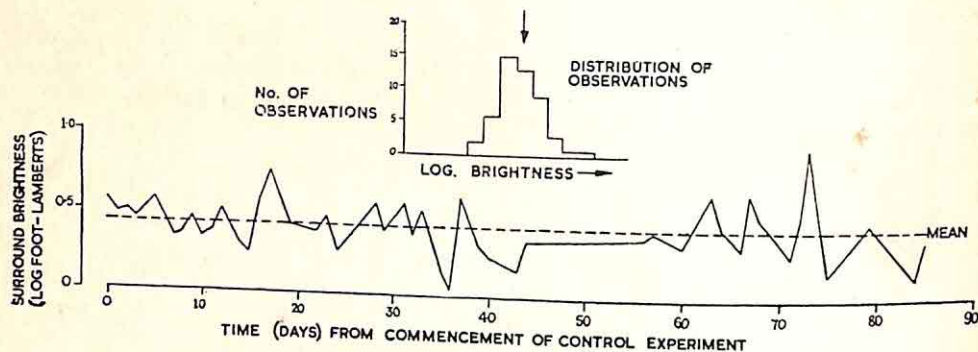


FIGURE 1

TYPICAL SET OF OBSERVATIONS OF A "CONTROL" EXPERIMENT. ASSESSMENT OF SURROUND BRIGHTNESS WITH WHICH GIVEN GLARE SOURCES GIVE "JUST UNCOMFORTABLE" GLARE

outside the limits twice the mean, and half the mean, more than once in twenty times. (This corresponds to a coefficient of variation of 0.42). Hence only once in twenty times would the setting for one criterion fall into the range of settings for the adjacent criterion. This procedure was followed in the subsequent studies, a few consistent observers being preferred to a larger number with lower consistency. As far as possible, staff who were associated with the results of the investigation were not themselves employed as observers. Fig. 1 is a typical plot for a consistent observer performing the control experiment. Fig. 2 shows a plot, under the same conditions,

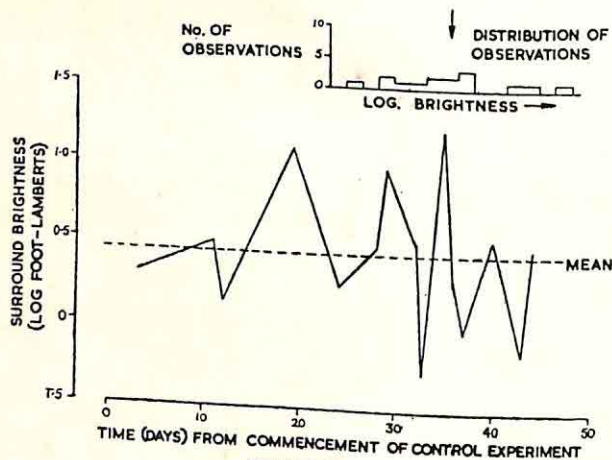


FIGURE 2

ASSESSMENT MADE BY INEXPERIENCED OBSERVER
(Compare with Figure 1)

of a less consistent one. Fig. 3 shows a set of results by a consistent observer, Fig. 4 by an inexperienced and less consistent one.

It was found that observers varied considerably in the amount of confidence they felt and showed in their settings, and that in general those with the less confidence showed the greater scatter. In some cases the individuals who showed uncertainty could be induced to "guess," and such guesses often resulted in a surprising consistency of setting. If such an individual was now shown that his performance was not as poor as he expected, his confidence was assisted, and in this sense he could be "trained" to become a "useful" observer.

Choice of Criteria: As the technique was extended to a wider range of problems, it was found that the value of an experiment could be increased by careful attention to the choice of criteria. In particular, the form of words used in the description of each criterion had to be one which was concretely and directly applicable to the observer's own experience. For example, the set of criteria "just visible," "moderately visible," "reasonably visible," "easily visible," would be less satisfactory than "just visible," "just recognizable," "pattern just identifiable," "pattern definitely identifiable." An experienced observer could work with the first set, allocating to

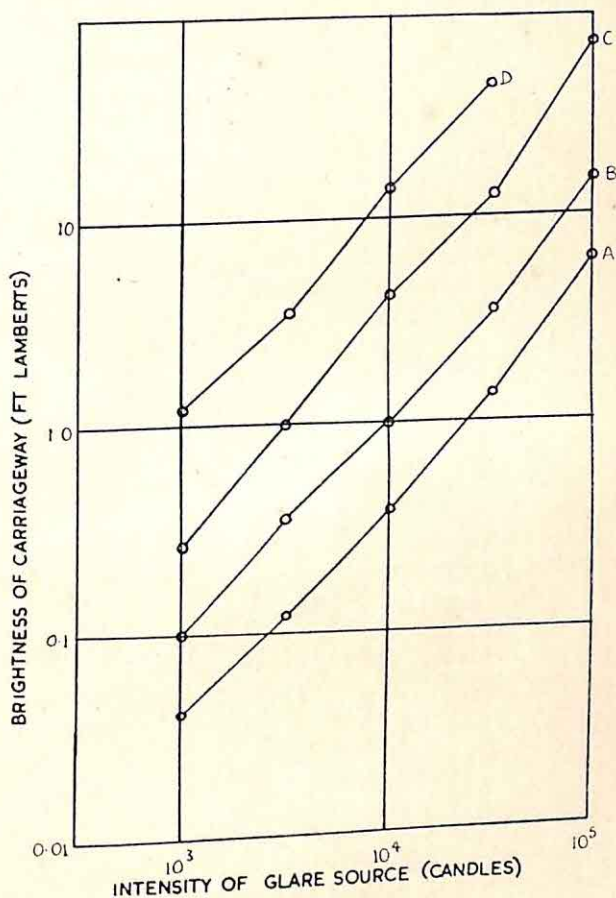


FIGURE 3
TYPICAL SET OF OBSERVATIONS SHOWING ASSESSMENT OF CARRIAGEWAY AND SURROUNDING BRIGHTNESS IN RELATION TO INTENSITY OF STREET-LIGHTING LANTERNS, FOR FOUR DEFINED CRITERIA OF GLARE (SHOWN BY A, B, C, D ON DIAGRAM)

each degree of "visibility" a mental assessment of the degree of identification of the pattern. But a less experienced one would not be able to make such an interpretation in a consistent way, and would be uncertain as to the exact meaning of "visible." An appropriate form of words did, in fact, generally emerge from the pilot experiment. If the criteria were not well defined, observers would express doubts as to what was required of them. More formal investigation of the effect of criteria used on consistency of judgment would be of considerable interest and value.

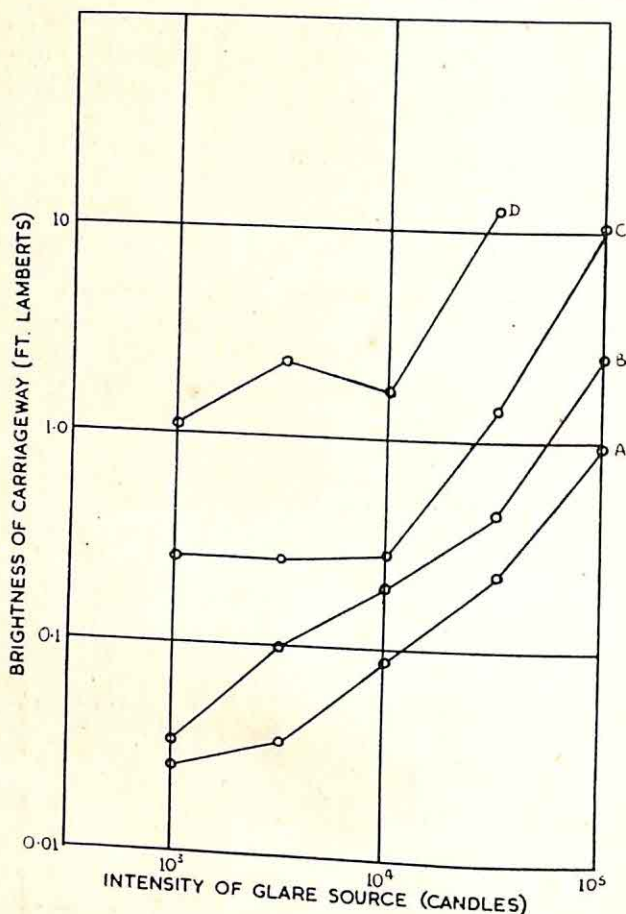


FIGURE 4
ASSESSMENTS BY INEXPERIENCED OBSERVER
(Compare with Figure 3)

III

EXPERIMENTAL PROCEDURE

The type of experimental procedure employed is perhaps best indicated by summarizing briefly the sequence of events in one experiment on radar trace visibility.

1. The observer enters the room, in which the general brightness level is about 0.1 foot lamberts. The lighting is indirect and free from glare and distracting features.

2. The recorder informs the observer that he is to make observations on the visibility of a radar trace on the face of the cathode-ray tube in front of which he is seated. The trace has a certain definite characteristic with which he is acquainted from previous observations under similar conditions.

3. A preliminary period of about five minutes is allowed during which the observer sits quietly and adapts to his surroundings. At the end of this time, the recorder sets the background brightness on the face of the cathode-ray tube to the first level to be studied.

4. The recorder then asks the observer to take control of the signal brightness. He is told that, as he raises the signal brightness, he will first become aware only of the existence and position of an area of higher brightness than the background, but will be unable to recognize its pattern.

At this point he is to call out "just visible." He is then to raise the brightness of the signal slowly until he reaches the point at which he judges that he can just recognize the characteristic pattern of two closely juxtaposed double traces. At this point he is to call "just recognizable." On further raising the brightness he is to call out "measurable" when he is quite sure that he can assess the relative lengths of the two double traces. Finally he is to raise the brightness to the lowest level at which in his opinion the signal is easily visible for continued working, and call "easily visible." Any further explanation of the criteria is given if required. When he fully understands the instructions, the observations commence.

5. The observer is free to take as long as he wishes. No assistance is given him unless he experiences difficulty or lack of confidence. The recorder logs the observations at each stage without comment. The observer has no indication of the values of his settings, and the controls are so arranged (e.g. by series resistance in a lamp brightness circuit) that the position of the control knob, or the angle through which it is turned, give no clue to the values reached.

6. The series of four criteria being completed at the first level of surround brightness, the observer has a short rest of about a minute while the next setting of surround brightness is set up by the recorder.

7. The process is repeated at the other surround brightness values.

A session will last from about ten minutes with an easy assessment, to about thirty minutes with a difficult one. The incidence of consciously felt fatigue is avoided.

IV

DISCUSSION

The validity of such a technique as that described must, in the first instance, be judged by the success of operations based on the conclusions to which its use may lead. On this score there is some evidence of its usefulness. Two independent teams of experimenters, working in one case on problems associated with the visibility of radar displays (Hopkinson, 1946), and in the other with a wide range of war-time visibility problems (Waldram and Stevens, 1946), found that reliable conclusions were reached from experiments which used the technique. It was from such work that decisions regarding agreed levels of war-time street-lighting were influenced. New designs of street lantern were evolved, too, on the basis of appraisals made by the technique, and these were found in practice to be more satisfactory than the old.

Another kind of check introduces a point of some theoretical interest. In certain cases it was possible to have small groups of observers, who had participated in the laboratory studies, make appraisals in actual streets, while at the same time measurements of the prevalent conditions could be made comparable with the laboratory set-up. In this case observers were asked to express the amount of discomfort experienced in terms of the criteria used in the laboratory experiment. It was found that observers who had been made familiar with the assessment of discomfort in terms of four fixed criteria were no longer reluctant to make *direct* assessment in terms of these criteria. The agreement between direct appraisal in actual streets, and indirect in a laboratory set-up similar as regards the physical variables involved was found to be good (Hopkinson, 1940, Table X).

Thirdly, it has been possible in some cases to check the appraisals against objective criteria of functional efficiency. In the work on the influence of brightness and size of radar echoes on ease of recognition, for instance, direct measurement was also made of percentage of accurate localizations. It was found that the variation between subjects in the mean percentage of accurate locations, corresponding to the settings of the variables for a given criterion of ease of recognition, was quite small, and amounted to about 5 per cent. This suggested that each criterion of ease of recognition did in fact correspond to a real standard of performance.

More recently, tests on the effect of illumination and contrast on ease of reading have been made in which the multiple criterion technique has been combined with measurements of speed of reading. The observers who were employed in the tests were selected for certain characteristics of their vision, and not for their intelligence or ability to operate the technique. The variation of their assessments might therefore be expected to be greater than that of the picked observers of the radar experiment. Table I summarizes some of the observations that were obtained.

It can be seen that "barely readable" and "reasonably readable" correspond to mean speeds of 0.4 and 0.7 of the maximum speed respectively. The coefficient of

TABLE I

VARIABILITY* OF MEASURED VALUES OF SPEED OF READING (EXPRESSED AS A RATIO OF MAXIMUM SPEED FOR EACH OBSERVER) FOR TWO FIXED CRITERIA OF EASE OF READING. (EACH OF 20 OBSERVERS GAVE ONE VALUE OF READING SPEED FOR EACH CRITERION OF EASE OF READING, UNDER EACH OF 12 CONDITIONS OF ILLUMINATION OR CONTRAST.)

Condition	Ease of Reading			
	Criterion B "Barely Readable"		Criterion C "Reasonably Readable"	
	Speed of Reading Mean of 20 Values	Variability about Mean	Speed of Reading Mean of 20 Values	Variability about Mean
1	0.35			
2	0.36	39.5%	0.69	17.9%
3	0.34	26.2%	0.67	21.9%
4	0.31	22.1%	0.71	18.4%
5	0.40	36.1%	0.62	22.3%
6	0.43	28.8%	0.73	17.4%
7	0.41	25.0%	0.76	18.0%
8	0.40	34.1%	0.67	19.6%
9	0.43	37.2%	0.70	20.8%
10	0.39	38.0%	0.75	14.7%
11	0.41	28.2%	0.68	12.5%
12	0.41	28.3%	0.71	14.2%
		23.1%	0.77	15.6%
Mean for all 12 conditions	0.39	—	0.70	—
Variability of the means for each condition about the mean for all conditions	10%	6.8%

$$* \text{Variability} = \frac{\text{standard deviation}}{\text{mean}} 100\%.$$

variation of the means for each of the 12 conditions, with respect to the mean of the whole of the 12 conditions is about 0.08. That of the individual values for each

observer about the mean for any one condition is about 0.25. These figures suggest that even observers inexperienced in the multiple criterion technique can, by a suitably designed experiment, give assessments of reasonable accuracy, which link with comparable performance figures.

It must be pointed out, however, that the particular coefficient of variation found in each task is dependent upon a number of factors. Among these we may distinguish the inherent difficulty of the judgment required, from the suitability of any given experimental design (in terms of variable chosen for control, formulation of criteria, etc.) to facilitate the making of the judgment. Thus, if we find a coefficient of variation of 0.05 on a search task, and one of 0.20 on a reading test, this difference may in part be due to an inescapable difference of inherent difficulty in the two tasks, and only in part to a corrigible difference in experimental design. Further work is needed to establish the relative magnitude of these factors so far as may be possible. It might then become possible to predict, to some extent, the consistency to be expected of any particular experiment. A high coefficient of variation should not of itself cause the rejection of an experiment, because it may still yield information of value, which cannot be obtained in any other way.

The scatter of the separate means for each observer about the mean of all these means is not easy of interpretation. Part of the variance may be attributable to random experimental errors common to all observers, while another part may be due to real differences in the psychological characteristics of different observers. Conditions which may be acceptable to one observer may not be tolerable to another more sensitive observer. It is not possible to determine whether or not the subjects have all interpreted the criteria in the same way. Direct experiment can penetrate only to a limited extent into the nature and magnitude of the sensory reactions of an observer.

V

ACKNOWLEDGMENTS

The work described in this paper was first undertaken when the author was with the Research Laboratories of The General Electric Company Ltd., Wembley. The recent work is part of the programme of the Building Research Board, and this paper is published by permission of the Director of Building Research, Department of Scientific and Industrial Research. The author gratefully acknowledges the help of his colleagues in the development of these investigations.

REFERENCES

1. HOLLADAY, L. L. (1926). *J. Opt. Soc. Amer.*, 12, 271.
2. HOPKINSON, R. G. (1940). *Trans. Illum. Eng. Soc. (London)*, 5, 1.
3. HOPKINSON, R. G. (1946). *J.I.E.E.*, 93, IIIA, 795.
4. HOPKINSON, R. G. (1949). *Trans. Illum. Eng. Soc. (London)*, 14.
5. PETHERBRIDGE, P. and HOPKINSON, R. G. (1950). *Trans. Illum. Eng. Soc. (London)*, 15, 39.
6. STILES, W. S. (1929). *Proc. Roy. Soc.*, 104B, 322.
7. WALDRAM, J. M. and STEVENS, W. R. (1946). *Trans. Illum. Eng. Soc. (London)*, Conventional Proceedings, 51.

(Received 5th August, 1949.)

THE INFLUENCE OF AN INTERPOLATED EXPERIENCE UPON RECOGNITION

BY

D. RUSSELL DAVIS and DURGANAND SINHA

(From the Psychological Laboratory, University of Cambridge)

It was demonstrated in a previous experiment that an experience interpolated between an original experience and its recall may bring about changes in the points of emphasis in the recall of the original experience. Moreover, details of the interpolated experience may be recalled as if they had formed part of the original experience. These results were taken to mean that two experiences of a related kind may become merged in memory into something akin to Bartlett's notion of an organized mass of past experiences. In the experiment here reported, the original experience was the hearing of a story, and the interpolated experience the seeing of a picture which illustrated part of the story. When in a recognition test subjects were asked to select from three alternatives (including the original) the one version which was "most like the original story," a proportion of them preferred to the *original story* a version which differed from the original by including a number of details from the picture. Asked about details, all the subjects tended to place details from the picture in the story, even if they had not been mentioned there. The results of the two experiments are thought to show that irreversible changes are brought about in the memory of an experience by subsequent experiences of a related kind.

I

INTRODUCTION

WE have recently reported an experiment in which the recall of one experience was shown to be modified by a second experience of a related kind (Davis and Sinha, 1950). This experiment was one of a series having the form:

Group	First Session: Original Experience	Second Session: Interpolated Experience	Third Session: Description, Recall or Recognition Test
1	Experience A	Experience B	Experience A
2	" A	—	" A
3	" A	Experience B	" B
4	—	" B	" B

In the experiment already reported, the material forming the original experience was a story, and that forming the interpolated experience a picture which illustrated more or less accurately a part of the story. These two experiences were shown to interact. Thus the perception and recall of the picture (as judged by the comparison of the results obtained in the third session from Groups 3 and 4) were governed by the attitudes induced by the story. These attitudes directed attention to, and prescribed the interpretation of, those details in the picture which played an important part in the theme of the story. Thus perceived, the picture influenced the recall of the story (as judged by the comparison of Groups 1 and 2). Seeing the picture brought about changes in the points of emphasis, favouring the recall of those aspects of the story which the picture illustrated. Details of the story not reinforced by the picture tended to be left out. Many picture details intruded into the reproductions of the story; that is, many details which were part of the interpolated, but not of the original experience, were recalled as if they had been part of the original.

These findings were taken as confirmation of Bartlett's notion that memories are retained as *schemata* or active organizations of related experiences, which are subject to modification or reconstruction as a result of new experiences of a relevant kind (Bartlett, 1932). Bartlett implied that the changes which memories undergo are irreversible, and that memories lose their separate identity and become incapable of recall in their original form. On the other hand, so far as they are pertinent, the findings argue against Freud's contention that in some way or other memories survive unchanged in the mind, as a rule rather than as an exception; thus Freud wrote in 1930, "... we have been inclined to the opposite view that nothing once formed in the mind could ever perish, that everything survives in some way or other, and is capable under certain conditions of being brought to light again, as, for instance, when regression extends back far enough." ... "The fact is that a survival of all the early stages alongside the final form is only possible in the mind." (Vide Freud, 1930, pp. 15-20.)

In contrast to Freud, experimental psychologists have in general asserted that in whatever form they are retained, memories become modified with lapse of time, as a result either of the intrinsic stresses arising from their configurational characteristics or of the extrinsic stresses resulting from antecedent or subsequent experiences. Indeed, many experimental studies have been carried out with the purpose of defining the changes to which memory traces or *engrams* are subject. Two types of data can readily be obtained: those from recall or reproduction tests, and those from recognition tests, these data being assumed to reveal changes in the engrams themselves. This assumption is valid only within limits, however, for it is often belied by the ancillary evidence to some extent; for instance, recognition tests sometimes yield results different from those yielded by recall tests; again, introspections often suggest that the material recalled does not fully represent the engram. Moreover, those who wish to insist upon the validity of Freud's theory of memory may treat the changes observed in recall or recognition tests, not as evidence of changes in the engrams, but merely as changes in the operation of the mechanism of repression.

Nevertheless, there are good reasons for the careful study of the results of recall and recognition tests, although they may have to be interpreted with caution. Accordingly we have examined the results of recognition tests in experiments of the above design, to discover whether they show the same degree of confusion between original and interpolated experiences as was shown by the recall tests described in our previous paper.

II

METHOD AND PROCEDURE

Of the several methods available, that of successive or repeated reproduction has been the most widely used in research on these problems. It is especially applicable when the autonomous changes due to intrinsic stresses are studied (e.g. Zangwill, 1937). Yet it has certain disadvantages in the study of extrinsic stresses, for in it there is no control or manipulation of the other related experiences from which these stresses are held to arise. Bartlett applied it in many of his most interesting experiments and was able to demonstrate in a convincing manner that recall was modified by rationalization and conventionalization, which represented the effects of antecedent experiences, but he was not able to show in any precise way how these effects came about. Zangwill (1939), too, used it in an experiment similar in some respects to the one to be reported below. Some of the disadvantages of this method have been avoided by the method of interpolated experience, which we have used.

In the experiment to be reported, the material forming the original and the interpolated experience was essentially the same as in the previous experiment. The original experience was the hearing of a shortened version of the story (about 500 words in this experiment) which was written for the purpose, and the interpolated experience was formed by Pieter Breughel's picture, "The Village Wedding." The story gave a new significance to the scene of festivity portrayed in this picture, its first part describing a feud between two peasant families which culminated in their reconciliation by the betrothal of the son of one family to the daughter of the other, and its main part a wedding feast disturbed by several untoward incidents.

This story was read out to forty university students, divided between Groups 1 and 2. The findings in Groups 3 and 4 are not pertinent to the present enquiry. Twelve days later half of them (Group 1) were shown individually a set of seven postcards, from which they were asked to pick out the one relevant to the story. All the postcards were coloured reproductions of pictures. Everyone selected "The Village Wedding," for reasons which are discussed in the previous paper, and was then given a short time to examine it. The postcards were not shown to the subjects of Group 2.

After an interval of from five to six weeks, all the subjects of Groups 1 and 2 were given (i) a recognition test of the story, and (ii) a questionnaire test. Comparison of the results in the two groups reveals the effects in Group 1 of the interpolated experience.

(i) In the recognition test, three versions of the story were read to each subject who was then asked to rank them in order of their similarity to the original story. The order in which the versions were read was controlled. Subjects were also asked to rate the certainty of their ranking, and to mention the factors which determined it. Finally, typed copies of the versions were given to them to read, and notes were made of their further comments.

Version A was the original story, unaltered.

Version B was similar to the original story in style, form and length. Its first part was unaltered, but the main part describing the wedding feast included several details present in the picture, but not in the original story; some details from the original story were omitted and this version provided a better account of the picture than did the original story.

Version C differed considerably from the original in detail, although similar in length and in theme.

(ii) Both questions and answers in the questionnaire test were written. The test comprised 35 questions, each asking whether a detail had been mentioned in the original story. An answer, "Yes," "No" or "I don't know," was invited to each. Three types of detail were asked about:

A details, present in the original story;

B details, not present in the original story, but present in the picture;

C details, present neither in the original story, nor in the picture.

The questionnaire test was applied after the recognition test. Some B details were contained in Version B, others were not. Some C details were contained in Version C.

III

RESULTS

(i) *Recognition test.*

Version C was rejected by all subjects. Eighteen of the subjects of Group 2 ranked Version A as most like the original; 11 of these stated it to be the original.

Two subjects of this group were doubtful whether to prefer Version A or B; after reading the versions for himself, one remained doubtful, and the other identified Version A as the original.

Five subjects of Group 1 ranked Version B as the most like the original; 14 preferred Version A, and of these, 10 identified it correctly as the original; one subject was very doubtful, regarding Version B as more like the original than Version A, except for one or two details; after reading the versions for himself, however, he preferred Version A.

Those subjects of Group 1 who ranked Version B as the most like, did so with less confidence than that with which the majority of the rankings were made. On the other hand, some of the subjects of Group 1 who preferred Version A were dissatisfied with this ranking; for instance, three such subjects remarked that details present in the original had been left out of Version A: "Some sort of reference to servants carrying food is absent in A." "In A villagers sitting on frail wooden benches have been left out." "There was something about music, though no music-benches were explicitly mentioned. That is also left out in A." All the details mentioned in these remarks are properly picture details which were not present in the original story. It was noted in this connection that subjects of both groups tended to regard Version A as less detailed and more condensed than had been the original, but it was only these three subjects of Group 1 who mentioned details which (they thought) had been left out.

That five subjects of Group 1 should have preferred Version B to A shows that recognition of an original experience may be affected by an interpolated experience. It is the more striking, since the details which were introduced into Version B from the picture were necessarily expressed in unfamiliar phrases. Indeed Version A was often preferred by subjects of Group 1 because the phrasing was familiar, e.g. subjects said of Version A: "Literary style the same—lots of sentences were familiar." "The tone of A struck me as identical with that of the original and every detail of description and action seemed familiar." "The gay chatter of the bride . . . their faces showed only a childish yearning for food as saliva dribbled from their mouths"—in fact, I was waiting for these phrases." Similarly, the unfamiliarity of phrasing in Version B was a main reason for its rejection by at least four members of Group 1. Again, several others of this group rejected Version B because it omitted details remembered in the original.

Selection or rejection of a version thus depended in many cases upon the presence or absence of a small number of details. For this reason the recognition test was less satisfactory than might have been supposed.

(ii) *Questionnaire test.*

The two groups did not differ in their answers to the questions concerning A details, about 70 per cent. of the questions receiving affirmative (correct) answers, about 20 per cent. negative (wrong) answers and about 10 per cent. "I don't know" answers. Similar results were obtained for C details. For B details, the following results were obtained:

	<i>Affirmative</i>	<i>Negative</i>	<i>I don't know</i>
Group 1	90	100	30
Group 2	47	136	37

(There were 20 subjects in each group, and 11 questions concerned B details.)

In the case of B details, affirmative answers were incorrect. That Group 1

should have given a significantly larger number (P being less than 0.01) of such answers than Group 2 indicates that details of an interpolated experience *intrude* into the memory of an original experience.

IV

DISCUSSION

Zangwill (1939) has reported a similar experiment, without the deliberate use of an interpolated experience. His subjects reproduced a prose passage three months after it had been read—a much longer time interval than in our experiment. This reproduction was used to construct a version including the salient inaccuracies in recall. Four or possibly five of his seven subjects expressed some difficulties in deciding in a recognition test, whether this version or the original story was the original and suggested that the original story should contain some of the inaccuracies present in the reconstructed version. Our results were similar, as was the experimental method, except in so far as we brought about the modification of the memories of the original experience experimentally, and Zangwill allowed the modification to come about with the passage of time.

The results of the present and the previous experiment allow the interpretation that an interpolated experience brings about changes in the recall or recognition of an original experience, accurate recall or recognition becoming impossible. It is not to be supposed, however, that there is complete compounding of the two experiences. On the contrary, they seem to retain their separate identities to some degree; some details, for instance, remain capable of correct recall or recognition as derived from the original experience, others as derived from the interpolated experience; only some details are confused, intruding from one experience into the other.

If those details which intruded from the interpolated into the original experience governed recognition of the original experience entirely, Version B would have been preferred invariably, or almost invariably, in our experiment. But this was not the case, as the reports of our subjects made clear; preference in the recognition test was determined by other factors in the case of many of the subjects, and depended upon factors which had not been influenced by the interpolated experience or had not been adequately controlled in the preparation of Version B. Again, modifications might well have been brought about in the memory of the original experience, which did not correspond to those introduced by the experimenter into Version B. The hypothesis that an interpolated experience brings about changes in the recognition of an original experience is not disproved by the preference for Version A shown by the majority of the subjects in Group 1., on the other hand, a hypothesis of independence of the two experiences would seem to be disproved by the preference by a minority of Version B.

The questionnaire test provided a more straightforward proof of a more restricted hypothesis that the memory of an original experience becomes modified by the intrusion of details from an interpolated experience. The results show that at least some of the intrusions which had been predicted came about.

Taken together, the results of the recall experiment reported previously and of this experiment indicated that, under the conditions obtaining, memories of related experiences do become merged to some degree, in a manner corresponding to Bartlett's notion of an "organized mass" of past experiences. There is no evidence immediately available to show that the merging is reversible. Probably it is not, any apparent reversal being due, not to a separation of the two experiences, but to a reconstruction by inference.

REFERENCES

1. BARTLETT, F. C. (1932). *Remembering*. Cambridge.
2. DAVIS, D. R. and SINHA, D. (1950). The effect of one experience upon the recall of another. *Quart. J. exp. Psychol.*, **2**, 43-52.
3. FREUD, S. (1930). *Civilization and its Discontents*, trans. Riviere, J. London.
4. ZANGWILL, O. L. (1937). An investigation into the relationship between the processes of reproducing and recognizing simple figures with special reference to Koffka's Trace Theory. *Brit. J. Psychol.*, **27**, 250-76.
5. ZANGWILL, O. L. (1939). Some relations between reproducing and recognizing prose material. *Brit. J. Psychol.*, **29**, 371-82.

(Received 20th February, 1950.)

MISCELLANEA

NOTE ON THE CORRELATION OF ERRORS AND PERFORMANCE IN A TEST

BY

J. W. WHITFIELD

(From the Medical Research Council Applied Psychology Research Unit, Cambridge)

In the common choice-response type of intelligence test, if an individual's intellectual ability fails him for a specific question he has the choice of omitting the question or of hazarding a guess. A similar choice is open in a number of experimental learning problems. By chance he will guess correctly in some cases, more often he will be in error. It is highly likely that individuals differ in their willingness to guess; this difference may or may not be related to intelligence or test sophistication, or such a relationship, if found, might be susceptible to change in test conditions.

A pure measure of "guessing" is out of the question, but apart from "catch" questions, the number of wrong answers given can be considered as an impure or indirect measure of guessing. There is no way of deciding which correct answers were in fact guessed. Thus the guessing evidence consists of the number of wrong answers and the number of questions unanswered. These two are, however, constrained by the total number of questions in the test and the total number of correct responses. It is obvious that an individual who correctly answers nearly all the test questions can demonstrate very little guessing, as it has been defined, whereas one with few correct answers can show many guessed answers.

Let us consider a test of " n " questions and a population N attempting the test. Let the mean number of questions answered correctly be C , and the corresponding incorrect (i.e. errors and unanswered) be I ($C + I = n$). Let the mean number of errors be E , and let the incorrect (or correct) answers be distributed about their mean with standard deviation σ . If there is no variation in guessing propensity amongst subjects, then for individuals who show x incorrect answers the errors will be distributed according to the binomial expansion $(p + q)^x$ where $p = \frac{E}{I}$.

Thus for the whole range of possible incorrect scores we have the following probability table:—

Number of Incorrect Answers i	P	Mean Number of Errors e	Variance of Error Distribution	Product Term (Incorrect \times Errors $= i e$)
0	P_0	0	0	0
1	P_1	p	$p(1-p)$	p
2	P_2	$2p$	$2p(1-p)$	$4p$
3	P_3	$3p$	$3p(1-p)$	$9p$
.
.
n	P_n	np	$np(1-p)$	n^2p

$$\begin{aligned} \Sigma P &= 1, \quad \Sigma P_i = 1, \quad \Sigma P_e = E = pI \\ \Sigma P_i e &= p\sigma^2 + pI^2 \quad \{\text{i.e. } p \Sigma P(i)^2\} \\ \Sigma P(e)^2 &= p^2 \{P_1 + 4P_2 + 9P_3 + \dots + n^2 P_n\} \\ &\quad + p \{P_1 + 2P_2 + 3P_3 + \dots + n P_n\} \\ &= p^2 \{\sigma^2 + I^2\} + pI - p^2 I \\ \therefore \text{Var } e &= p^2 \sigma^2 + pI(1-p). \end{aligned}$$

This expected variance itself provides data upon which to consider test behaviour. If individual differences in proneness to guess exist, and are uncorrelated with test score, the observed variance of the error distribution will exceed that calculated by the above formula. If, in addition, there is a correlation in the direction of the poorer the test score, the greater the likelihood of guessing, then the observed variance is considerably increased. If, however, the correlation is in the reverse direction, i.e. the better the test score the greater the number of guesses, then the variance will tend to be reduced, particularly if the high scores on the test are fairly near to the maximum possible performance.

The expected regression of e upon i has slope p , and the observed and expected regressions can be drawn to aid inspection. Alternatively the observed and expected product moment correlation between i and e can be calculated.

$$\Sigma P.i.e. \text{ is } p\sigma^2 + pI^2.$$

The expected value of r is therefore—

$$\begin{aligned} & \frac{p\sigma^2 + pI^2 - IpI}{\sigma\sqrt{p^2\sigma^2 + pI(1-p)}} \\ = & \frac{p\sigma}{\sqrt{p^2\sigma^2 + pI(1-p)}} \end{aligned}$$

It will be seen that the expected correlation between the number of incorrect and the number of errors will always be positive, tending towards unity as p tends towards 0 or 1, and, for given values of I and σ it will be at a minimum when $p = \frac{1}{2}$.

The relation between the observed $i e$ correlation, and that expected on this null hypothesis, taken in conjunction with the observations on the variance provides information on the behaviour of individuals in the intelligence test situation.

(Manuscript received April, 1950.)

A SIMPLE CLASS-ROOM CHRONOSCOPE

BY

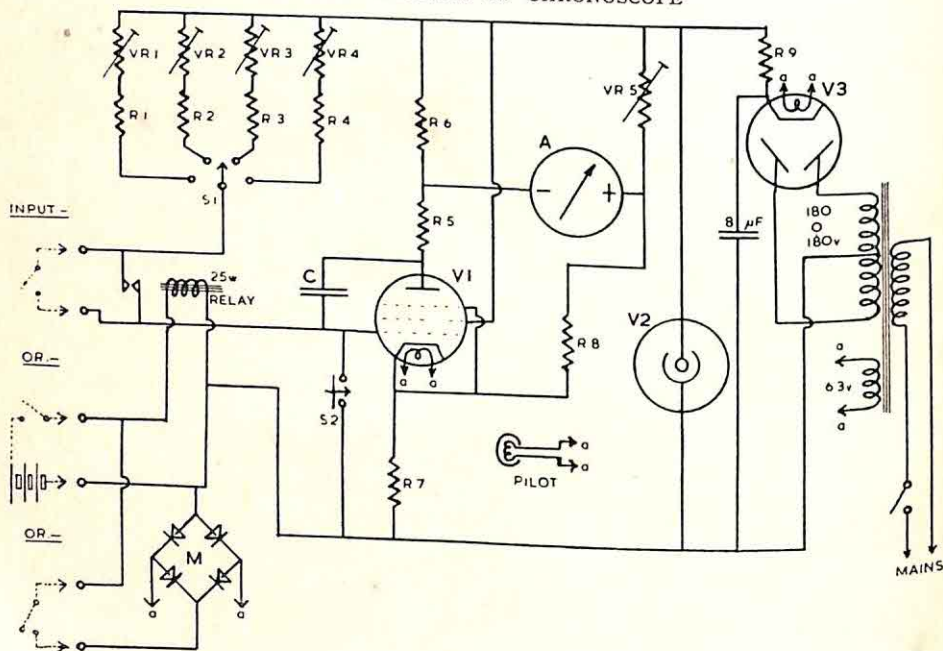
A. CARPENTER

(From the Medical Research Council Applied Psychology Research Unit, Cambridge)

THIS note describes an electric chronoscope which is simple and cheap to make, and which while not pretending to a high order of accuracy, is adequate for the majority of classrooms, and many research purposes.

The chronoscope described here was made in response to a request for an apparatus simple enough to operate and consistent enough in its readings for routine use by a class of students; yet sufficiently simple in design and construction, for a number of them to be made quickly. It was desired to measure the time for which an electrical circuit was closed, over a range of approximately 0.1 to 5 seconds, and it was decided that an accuracy of the order of 2 per cent. of the full-scale reading would be adequate. The choice was between a mechanical system operated by a constant speed motor, relying upon the accuracy of the frequency of the electric mains supply, and using some form of magnetic clutch to engage the scale pointers with the drive, as in the Hipp, and an

FIGURE 1
CIRCUIT DIAGRAM OF CHRONOSCOPE



electric system relying upon the charging of a condenser from a constant voltage source. The latter was decided upon for two reasons: (a) for consistent operation, a mechanical system would require to be constructed with considerable care and skill, and would therefore be costly, and (b) recent developments in valve-operated electrical circuits have made the latter method practicable and simple. Using standard small radio components it is possible to complete this instrument within two days, and for a cost of the order of £5, plus the cost of a suitable meter.

The circuit is given in Figure 1. It will be seen that the condenser, C, is charged from a voltage derived from the mains in the usual way, and stabilised with a gas discharge valve V_2 . Four ranges are provided, corresponding to full-scale readings of 0.5, 1, 2 and 5

seconds, and are selected by the four-position switch S_1 , which selects the resistor through which the condenser is charged. These resistors each consist of standard 1 watt carbon components R_{4-1} , in series with pre-set carbon potentiometers, VR_{4-1} . The accurate setting of these pre-set controls determines the absolute accuracy of the readings obtained, but for those purposes for which comparative readings are adequate, it is unnecessary to set these controls accurately. The condenser, C , is placed between the anode and control grid of the valve V_1 , thus making use of the Miller effect to obtain linearity of charging rate. With the valve and component values given the grid voltage actually changes by about 2 volts, so that at full-scale reading the rate of charge is just more than 1 per cent. less than the initial rate.

The insulation of the grid circuit of V_1 is of the highest importance in avoiding drift of the meter reading when the operating circuit is open. For this reason the input terminals are mounted on an insulating panel of Paxolin, and also, to enable the instrument to be connected to external apparatus whose insulation may be imperfect, a high-speed relay, operated either from the valve low tension supply, or from an external source of some 6 volts, is included. The 4 microfarad condenser, C , must be chosen to have as low a leakage as possible. In the instrument made the drift was unreadable over a period of five minutes.

The valve used as V_1 is not critical. A high mutual conductance is an advantage, and a top grid connection is desirable to minimise leakage.

The meter, which reads the anode current of V_1 , determines the accuracy of the complete instrument. In the first model made this was a panel mounting instrument with a $2\frac{1}{2}$ in. scale, the whole being made as one unit, but in later copies larger meters have been used, connected externally. There is no reason why the meter should not be several yards distant from the rest of the instrument. Zero is adjusted by VR_5 , the minimum valve current to be backed off by this means is about 0.5 mA.

Switch S_2 is a push-button switch which on closure restores the meter reading to zero by allowing C to charge almost to the full H.T. potential, through R_5 .

Possible sources of inaccuracy in this instrument are: (a) faulty setting of VR_{1-4} will produce absolute error of calibration; (b) faulty zero setting of VR_5 will produce a constant error; (c) if used on low ranges the operating time of the relay may introduce a constant error; (d) non-linearity due to exponential charging of C . As mentioned before, this is about $1\frac{1}{2}$ per cent. in this instrument; (e) leakage causes slow drift of the meter reading. The meter should be returned to zero just before taking each reading, and read without undue loss of time. In this instrument, drift was negligible. Finally, attempts to assess unreliability due to random causes have shown that readings can be relied upon to be repeatable to within the error involved in reading the meter scale—less than 1 per cent. of the full scale reading.

This circuit involves no new principles or methods. Fuller information will be found in the references given.

REFERENCES

- SOWERBY. "Interval timers." *Wireless World*, LV, 1, January, 1949, pp. 23-4.
 GRAY, J. W. "Integrating Circuits." Massachusetts Institute of Technology, Radiation Laboratory Handbook Series, Vol. 21, on Electronic Instruments, pp. 78-82.
 McGraw Hill, 1948.

COMPONENT VALUES

V_1	—	SP61	R_1	—	150 K.
V_2	—	VR_{150}	R_2	—	250 K.
V_3	—	6 X 5	R_3	—	500 K.
A	—	0.5 mA, 25 ω .	R_4	—	1 M.
C	—	4 microfarad high insulation	R_5	—	15 K.
M	—	0.25 A. metal rectifier	R_6	—	220 ω .
VR_1	—	50 K.	R_7	—	270 ω .
VR_2	—	100 K.	R_8	—	15 K.
VR_3	—	500 K.	R_9	—	2.5 K.
VR_4	—	1 M.	VR_5	—	30 ω .

BOOK REVIEW

The Organization of Behaviour: a Neuropsychological Theory. By D. O. HEBB. New York: J. Wiley & Sons; London: Chapman & Hall, Ltd. Pp. xix + 335. 32s. net.

In this lively and intelligent book, Dr. Hebb has tried to develop some broad principles of nervous organization which might have explanatory value in psychology. His aim throughout is to find conceptions "... which will be valid physiologically and at the same time 'molar' enough to be useful in the analysis of behaviour" (p. 11). If the task is difficult, Hebb has none the less made an original and exciting beginning. In the stagnant atmosphere of contemporary psychological discussion this little book comes like a breath of fresh air.

The earlier chapters are concerned in part to demolish current formulations and in part to suggest more acceptable alternatives from the standpoint of neurophysiology. First, the notion that behaviour is (or can ever be) completely controlled by peripheral sensory processes is decisively rejected—and with it all varieties of learning theory narrowly based on the reflex-arc concept. Second, the concepts of equipotentiality and "field dynamics" are examined and found wanting. As Hebb justly observes, Gestalt theories of visual organization have not avoided the difficulties presented by the anatomy of the visual cortex (p. 56). In consequence, he finds himself impelled to seek new conceptions of the central mechanisms governing behaviour more firmly based on the anatomy and physiology of the nervous system. Although the critical discussion is always informed and often acute, it is the constructive aspects of Hebb's work which merit the closest attention. These may accordingly be examined.

Hebb's positive contribution to psychological theory is given by two new conceptions, that of the *cell-assembly* and that of the *phase-sequence*. The cell-assembly is conceived as a specific and localized system of neurones firing in accordance with a sequence imposed by earlier patterns of excitation. It is held to comprise a large number of alternative pathways with equivalent functions and is thus endowed with a measure of equipotentiality (p. 74). In the course of individual experience, cell-assemblies are held to become organized into larger functional units according to a principle of phase-sequence. This notion is never very adequately defined in general terms but its meaning becomes fairly clear in the course of the discussion of perception (pp. 79-106). In the visual perception of form, Hebb argues, the trend of recent evidence suggests that the essential mechanisms are in large part acquired. Recognition of form and pattern, he suggests, may be traced to the integration of a temporal sequence of oculo-motor adjustments (successive shifts of fixation) with the corresponding sequence of visual data. There is thus built up a specific sequence of cortical events each with its central motor component. When perception of a familiar visual form is demanded, it is supposed that a series of cell-assemblies fires off according to a particular phase-sequence determined by previous activity. Even when recognition becomes automatic, and takes place without overt eye-movement, its central basis must still be envisaged in terms of the sequential activity of a sensori-motor "schema." Hebb suggests further that manual and locomotor habits are to be analysed in the same way but no detailed illustrations are given. In general, it is argued that the concept of phase-sequence may provide a theoretical groundwork for dealing with all temporally organized nervous processes, not excluding those governing the train of thought (p. 100).

This bald statement of Hebb's theories in no sense does justice to the subtlety of the argument or to the wealth of pertinent data drawn from experimental and clinical sources. At the same time, the author is careful to point out that his notion of the cell-assembly differs little in principle from earlier conceptions. An unkind critic might even say that the whole argument is little more than the Jamesian theory of habit furnished up to meet the more exacting demands of present-day neurophysiology. But this would be unfair as well as unkind. Hebb's treatment, speculative though it is, does serve to throw emphasis upon important features of perceptual reaction signally neglected in current psychological theory. In particular, one may mention the importance attached to motor components in the general development of perception and the insistence upon a genetic approach to visual problems. These aspects of Hebb's work may well give rise to a much-needed change of attitude in many departments of experimental psychology.

The later chapters of this book are devoted for the most part to a discussion of learning, motivation and intelligence. In his treatment of motivation, especially, Hebb makes a most praiseworthy attempt to break away from the fixed ideas of contemporary American learning theory. It is not unfair to state that the latter envisages all modification as a quasi-conditioning process governed by those mythical phantasms—the drives. For Hebb, on the other hand, motivation is defined in terms of the organization, direction and persistence of phase-sequences (p. 181). That is to say, drive is an inherent property of the neural system upon which execution of a habit or skill depends. Although this notion needs working out in much greater detail, it is a notable advance in theoretical approach. There is also a sensible, if less original, treatment of learning in its comparative aspects and a good discussion of instinct. Hebb takes the view that instinct should not be regarded as fundamentally distinct from what is usually called intelligence. On the contrary, "... it is intelligence, or insight, that is innately limited in variety" (p. 169).

The book concludes with an entertaining discussion of emotion and a brief excursus into psychopathology (which includes two clinical histories of spontaneous neurosis in chimpanzees). There are also some shrewd and timely comments on intellectual deterioration and its problematical relations to the site and extent of cerebral lesion. Throughout, the emphasis is physiological but the author is refreshingly free from the limitations of doctrinaire behaviourism. He is interested in ordinary experience and ordinary conduct and is in search of the nervous mechanisms by which they are sustained and controlled. If he has failed to find them, at least he is steering in the right direction.

It is clear from this book that its author has made a sensitive and intelligent response to the major intellectual currents of his time. Although he clearly pins his ultimate faith on neurophysiology, he has learnt much from clinical neurology and neurosurgery, experimental and comparative psychology, and shrewd observations of animals and man in daily life. The only intellectual influence which has completely passed him by is that of Freud. In many, of course, this would be an advantage but in Hebb's case it is something of a limitation. His treatment of emotion, and indeed psychopathology in general, would assuredly gain from a determined attempt to come to grips with the Freudian position. Apart from this qualification (which may well be personal to the present reviewer), there is little in the book to which exception might be taken. One is struck, it is true, by a certain naïveté and by a tendency at times for enthusiasm to outrun discretion. This is most clearly shown in the overwhelming importance attached to von Senden's study of the restoration of vision in operated cases of congenital cataract. Much of the material assembled by von Senden is generally regarded as somewhat suspect and unsuitable as a basis for broad theoretical generalizations. One may comment, too, that the organization of the book is somewhat slap-dash and at times the argument becomes extremely difficult to follow. This may deter readers accustomed to the ordered "phase-sequence" of a scientific text. None the less, these are faults inseparable from constructive endeavour and merit the reader's indulgence. Hebb's book reflects a most creditable attempt to subdue the nervous system conceived as the instrument of behaviour, and to state points of view which at all events do not openly conflict with the facts disclosed by modern neurophysiology. In the present unhappy state of physiological psychology this is no mean achievement.

O. L. Z.

SOME RECENT PUBLICATIONS BY MEMBERS OF THE EXPERIMENTAL PSYCHOLOGY GROUP

- The Framework of Human Behaviour.* By J. M. Blackburn. London: Kegan Paul, Trench, Trubner & Co., Ltd., 1947. Pp. viii + 158. 12s. 6d. net.
- Dimensions of Personality: a Record of Research.* By H. J. Eysenck. London: Kegan Paul, Trench, Trubner & Co., Ltd., 1947. Pp. xl + 308. 25s. net.
- Pilot Error: some Laboratory Experiments.* By D. Russell Davis. Air Ministry (A.P.3139A). London: His Majesty's Stationery Office, 1948. Pp. 39. 9d. net.
- Researches on the Measurement of Human Performance.* By N. H. Mackworth. Medical Research Council Special Report Series No. 268. London: His Majesty's Stationery Office, 1950. Pp. 156. 4s. od. net.



✓



